ESSAYS ON TESTING STRUCTURAL CHANGES AND CONSTANT CONDITIONAL DEPENDENCE VIA THE FOURIER TRANSFORM

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Zhonghao Fu

May 2017

ESSAYS ON TESTING STRUCTURAL CHANGES AND CONSTANT CONDITIONAL DEPENDENCE VIA THE FOURIER TRANSFORM

Zhonghao Fu, Ph.D.
Cornell University 2017

This dissertation consists of three essays on testing structural changes and constant conditional dependence via the Fourier transform.

The first essay, "A Model-free Consistent Test for Structural Change in Regression Possibly with Endogeneity", proposes a consistent test for structural change in a nonparametric times series regression model based on the Fourier transform. It is well known that structural instability leads to misleading inference and imprecise prediction of stationary time series models. I propose a model-free consistent test for structural change in regression by testing the instability of the Fourier transform of data. This novel approach avoids smoothed nonparametric estimation of the unknown regression function and so is free of the "curse of dimensionality" problem, especially when the dimension of regressors is high. As a result, the proposed test is asymptotically more powerful against a class of local alternatives than Vogt's (2015) nonparametric test for structural changes, which is the only consistent test for structural changes in a nonparametric regression model in the existing literature. The nonparametric tests of Hidalgo (1995) and Su and Xiao (2008) are asymptotically more powerful than the proposed test against certain smooth local alternatives, but they are not consistent tests and are asymptotically less powerful against a class of non-smooth local alternatives. Unlike the existing literature, I allow for endogenous and discrete regressors. By using a proper choice of weighting functions

for the transform parameters in the Fourier transform, I avoid numerical integration so that the test statistic is easy to compute. The test statistic has a convenient asymptotic $N(0, 1)$ distribution under the null hypothesis of no structural change and is consistent against a large class of smooth structural changes as well as abrupt structural breaks with unknown break dates. A Monte Carlo study and an empirical application show that the test performs reasonably well in finite samples.

In the second essay titled "Consistent Testing for Structural Change in Time Series Regression Models via the Fourier Transform", I focus on testing structural changes in a linear time series regression model via the Discrete Fourier Transform (DFT). The intuition is straightforward: if the true model parameters are time-varying, then the conventional estimation methods like OLS or 2SLS will fail to estimate the unknown parameters consistently. The estimated residuals will contain the time-varying local feature of model parameters. The Discrete Fourier Transform of estimated residuals will contain this information and reveal it in the frequency domain. One can then infer the existence of structural changes regardless of whether they are smooth or abrupt. Compared to the existing consistent tests for structural change, my test avoids smoothed nonparametric estimation of the unknown time-varying parameters. The rate of the local alternatives that our test can detect is $T^{-1/2}$. Furthermore, my test is robust to unknown structural change in explanatory variables and instrumental variables, which makes the test widely applicable, especially in macroeconomic models. Simulation studies demonstrate its good finite sample performance. I apply my test to examine the stability of the hybrid New Keynesian Phillips Curve and find evidence of structural changes in 1980 to 2001 which is treated as a stable period by Zhang et al. (2008) and Hall et al. (2012).

The third essay, "Testing Constancy of Conditional Joint Dependence", proposes an omnibus test for the constancy of conditional joint dependence on some state variables. The test statistic is constructed by comparing the generalized conditional covariance function and the generalized unconditional covariance function. It detects if the dependence strength between any two random variables varies with a certain factor that we are interested in. I show that by using a special weighting function proposed by Székely et al. (2007), the test statistic can be easily computed without using numerical simulation. Also by nonparametric regression, I show that the test statistic is both computationally and asymptotically invariant to possible high dimensional data. Furthermore, the test statistic is asymptotically pivotal and follows a convenient asymptotic $N(0, 1)$ distribution under the null hypothesis. I also show that the test statistic can apply to many other testing frameworks with proper transformation. A simulation study shows that the test works well in finite samples.

BIOGRAPHICAL SKETCH

Zhonghao Fu was born in Qiqihaer, China in July 1986. He studied Political Science at Renmin University of China and earned his B.A. degree in Law in July 2009. He was admitted to Cornell Institute for Public Affairs and earned his M.P.A. in May 2011. He continued his graduate studies in economics in the Department of Economics at Cornell University and will earn his Ph.D. degree in May 2017.

This document is dedicated to my wife Yunxuan Zhang, my daughter Linyi Fu and my parents, Hanjun Fu and Lijun Wang.

ACKNOWLEDGEMENTS

I am deeply indebted to my advisor Prof. Yongmiao Hong for his continued support during my Ph.D. studies. His patience, motivation, and excellent guidance helped me in all the time of research and writing of this thesis.

I would like to express my deepest gratitude to Prof. Nicholas M. Kiefer and Prof. G. Andrew Karolyi, for their helpful discussions and encouragement as well as serving on my committee.

I also thank Xin Wang, Prof. Hong's wife. She treated me like her son and gave me the guidance and encouragement when I need them the most. This thesis could not have been completed without her support.

Last but not least, I want to thank my wife, my daughter, and my parents for their love and support. My wife and daughter make my life delightful. My parents have supported me firmly since the beginning. They are my source of strength during my job market period.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 A Model-free Consistent Test for Structural Change in Regression Possibly with Endogeneity

Nonlinearity often exists in economic time series data and various nonlinear time series models have been proposed to capture different forms of nonlinearities. Examples include the threshold model by Tong (1983), the smooth transition model by Teräsvirta (1994), the Markov switching model by Hamilton (1989), the artificial neural network model by White (1989), the functional coefficient model by Cai, Fan, and Yao (2000), and the nonlinear factor model by Bai and Ng (2008). These nonlinear models usually produce reasonable in-sample fit, but they often perform poorly in out-of-sample prediction; see (e.g.) De Gooijer and Kumar (1992), Clements, Franses, and Swanson (2004), González-Rivera and Lee (2009). Teräsvirta, Tjøstheim, and Granger (2010) attribute the predictive failure of nonlinear time series models to the structural change of the nonlinear features detected in-sample. They argue that only when the nonlinearity documented in-sample also exists in the forecasting period will the prediction by the nonlinear time series model outperform a linear time series model. However, the nonlinearity detected in-sample can change or even disappear out-of-sample, causing misleading inference and imprecise prediction.

Economic relationships may suffer from structural changes due to changing economic environments, e.g., shocks, new policies, preference changes, technology progress, etc. When the time span is long, it is likely that a time series econometric model becomes unstable over time. Many studies have investigated the impact of structural changes on modeling, inference, and prediction; see (e.g.) Goyal and Welch (2008) and Stock and Watson (2009). For any nonlinearity detected in-sample, there exist two possibilities: either the data generating process (DGP) is nonlinear and time-invariant, or the DGP changes over time. Both of these can generate a similar nonlinear pattern in the data, although their implications on modeling, inference, and prediction are different. Therefore it is important to detect whether the underlying economic relationship is time-invariant before using a suitable stationary time series econometric model. This chapter proposes a model-free consistent test for structural changes in regression so that it can be used to distinguish structural change from nonlinearity or model misspecification, among many other things.

Testing instability of parameters in a time series parametric regression model has drawn much attention in the literature. Often a linear time series regression model is considered. Chow (1960), Andrews and Fair (1988), Andrews (1993), Andrews and Ploberger (1994), Bai and Perron (1998, 2003), Hansen (2001), Elliott and Müller (2006), and Perron (2006) develop tests for parameter constancy against abrupt structural breaks, with known or unknown breakpoints. Lin and Teräsvirta (1994), Cai (2007), Chen and Hong (2012), Kristensen (2012), Zhang and Wu (2012), and Cai, Wang, and Wang (2014) test parameter constancy against smooth structural changes. All these tests are based on the assumption that the parametric regression model is correctly specified, and regressors are exogenous. When endogenous regressors are present, Hall, Han, and Boldea (2012), Perron and Yamamoto (2014, 2015), and Chen (2015) discuss how to model and detect structural changes of parameters in a linear time series regression model.

Tests based on a parametric framework are valid only when the parametric model is correctly specified. However, economic theory usually does not tell any concrete functional form for the regression function. When the parametric regression model is misspecified, a rejection of stability may be caused by model misspecification, rather than structural change. To avoid this drawback, we will use a nonparametric approach that does not assume any restrictive functional form for the underlying regression function. We emphasize that Hidalgo (1995) has pioneered to propose a nonparametric conditional moment test for structural change in regression without having to specify the unknown regression function. Su and Xiao (2008) propose CUSUM-type tests for structural change in a nonparametric time series regression model that allows nonstationary covariates. Su and White (2010) consider testing structural change in a partially linear regression model. Vogt (2015) also proposes a nonparametric test for structural change in regression by checking whether the shape of regression function is stable over time, which is to our knowledge, the only consistent test for structural changes in a nonparametric regression model in the existing literature. All these tests can detect instability of the unknown regression function, free of model misspecification. Because existing tests in the literature are based on smoothed nonparametric estimation of the regression function, they suffer from the notorious "curse of dimensionality" problem at least in finite samples, especially when the dimension of regressors is high. Also, they restrict regressors to be exogenous and continuous, and require the use of a higher order kernel. When covariates are endogenous or discrete, no existing test is available to test instability of the unknown regression function.

This chapter contributes to the literature on testing instability of an unknown regression function in several directions. First, unlike the existing approaches which involve nonparametric estimation of the unknown regression function, we use a Fourier transform approach that avoids nonparametric estimation of the regression function, which makes our approach free of the "curse of dimensionality" problem. This is achieved by testing the time-varying property of the Fourier transform of the unknown regression function. Our nonparametric estimation of the time-varying Fourier transform only involves smoothing over time, not smoothing over regressors which could be of a high dimension. As a result, our test is asymptotically more efficient than Vogt's (2015) consistent test for structural change under a class of local alternatives. Although our test is asymptotically less powerful than the nonparametric tests of Hidalgo (1995) and Su and Xiao (2008) under certain smooth local alternatives, it is a consistent test and is asymptotically more powerful under a class of non-smooth local alternatives. Furthermore, since we avoid nonparametric estimation of the unknown regression function, we allow regressors to be continuous, discrete or a mixture of both. In contrast, Hidalgo (1995), Su and Xiao (2008), and Vogt (2015) all require regressors to be continuous and impose certain smoothness conditions on their density function to obtain a better convergence rate for nonparametric regression estimation.

Second, we allow for endogenous covariates. In contrast, all existing nonparametric tests assume exogenous covariates. Nonparametric regression with endogeneity has drawn increasing attention in the literature; see, e.g., Newey, Powell and Vella (1999), Ai and Chen (2003), Newey and Powell (2003), Hall and Horowitz (2005), Blundell, Chen, and Kristensen (2007), Horowitz (2011), Darolles, Fan, Florens, and Renault (2011). By considering the Fourier transform of the unknown regression function, we avoid tackling the "ill-posed" inverse problem in nonparametric instrumental variable estimation. Our test can serve as a pre-test for instability of the unknown regression function before using any nonparametric instrument variable estimation. This greatly expands the literature since all existing nonparametric tests are only applicable to a regression framework with exogeneity. There are tests that allow for endogenous covariates in the literature, but they are restricted to a linear regression model (e.g., Chen, 2015).

Third, to ensure the consistency of our test against a large class of smooth structural changes as well as abrupt structural breaks, we have to integrate out a transform parameter vector whose dimension is the same as that of covariates or instruments, which is computationally challenging when the dimension of covariates is high. However, the computational burden can be greatly alleviated by using a proper weighing function for the transform parameter. We discuss weighting functions that avoid numerical integration. An example is the joint standard normal density function used by Hong, Wang, and Wang (2014) in testing strict stationarity. Finally, our test statistic is asymptotically pivotal, and has a convenient asymptotic null $N(0, 1)$ distribution.

1.2 Consistent Testing for Structural Change in Time Series Regression Models via the Fourier Transform

Testing for structural change in a time series regression model has drawn much attention in the literature, ever since Chow (1960) and Quant (1960). A time series dynamics usually suffers from abrupt structural breaks or smooth structural changes due to changing economic environments, e.g., shocks, policy shifts, technology progress, preference changes, etc. Many empirical studies have confirmed the prevalence of structural instability in financial and macroeconomic time series. For example, Welch and Goyal (2008) show that the in-sample significant predictability of most financial and macroeconomic variables fails to yield better out-of-sample forecasts of the U.S. equity premium than the simple historical mean equity returns. One possible reason is the instability of model parameters and it is confirmed by Chen and Hong (2012). In labor economics, Hansen (2001) finds strong evidence of a structural break in labor productivity between 1992 and 1996, and weaker evidence of a structural break in the 1960s and the early 1980s. In macroeconomics, Stock and Watson (1996) find substantial instability in 76 representative US monthly post-war macroeconomic time series. Zhang et al. (2008) show that the instability of parameters in the New Keynesian Phillips Curve results in conflicting conclusions about the key determinant of short-run inflation dynamics. Therefore, detecting the existence of structural change is crucial for econometric modeling and inference.

Although most existing studies focus on dealing with abrupt structural breaks (e.g., see Perron, 2006), estimation and testing with smooth structural changes have drawn increasing attention. Intuitively, it is quite likely that economic agents digest and react to shocks such as new policies and unexpected income in a gradual manner. Even when the change is abrupt at the individual level, it is likely to behave as a smooth change at the aggregate level. As a result, the parameters in a time series model usually change smoothly over time rather than shifts abruptly. Smooth structural changes can be modeled parametrically. One example is the Smooth Transition Regression (STR) model developed by Lin and Teräsvirta (1994), where a particular parametric function is chosen to model the time-varying parameters. However, economic theories usually provide no information on how parameters should evolve over time. To avoid misspecification, a nonparametric time-varying parameter model was introduced by Robinson (1989, 1991) and further studied by Orbe et al. (2000, 2005), Cai (2007), Chen and Hong (2012), Kristensen (2012), Zhang and Wu (2012), Cai et al. (2015), and Xu (2015). Among many others, Chen and Hong (2012) propose a generalized Hausman's test for both smooth structural changes and abrupt structural breaks in a linear time series regression model. Kristensen (2012) considers estimation and testing in both mean and variance in a time series dynamics. Cai et al. (2015) test the instability of model parameter when the covariates follow a unit root process. Xu (2015) constructs a CUSUM-type test for smooth structural changes in regression coefficient when the variance is time-varying. Although the tests mentioned above can detect the instability of parameters of unknown forms in a linear time series model, they are restricted in a conditional mean framework. When endogenous covariates are present, their tests are not applicable.

Several recent works have considered estimation and testing in a linear time series model with endogenous covariates. Hall et al. (2012) extend Bai and Perron's (1998) approach to estimation and testing for abrupt breaks of linear models with endogeneity using the Two-Stage Least Squares (2SLS). However, their approach requires firstly identifying the structural breaks in the reduced form. If there exist too many breaks or smooth structural changes in the reduced form, then their approach is not applicable. Perron and Yamamoto (2014) consider estimation and testing based on the 2SLS by extending Perron and Qu (2006). Perron and Yamamoto (2015) propose a testing method for abrupt structural breaks using OLS rather than 2SLS. They show the OLS-based test is more powerful than tests based on the 2SLS in most cases. However, the inference is restricted to the dates and magnitudes of breaks in the reduced form, and their test is not consistent against local alternatives of certain directions. Unlike the studies that focus on abrupt structural breaks rather than smooth structural changes, Chen (2015) proposes a Two-Stage Local Linear (2SLL) method for estimation and testing when the unknown parameters exhibit smooth change. However, the test statistic is computed via smoothed nonparametric estimation of the unknown possibly time-varying parameter which requires choosing bandwidths for both the structural function and the first stage reduced form. Furthermore, certain smoothness condition is needed to ensure the consistency of smoothed nonparametric estimation, which is restrictive since we usually have no prior knowledge about the property of unknown parameters over time. Moreover, the rate of alternatives that Chen (2015) can detect is $T^{-1/2}h^{-1/4}$, where $h \to 0$ is a bandwidth. This is slower than the parametric rate $T^{-1/2}$.

In this chapter, we propose a novel test for structural change in a time series model via the Fourier transform. Unlike the existing tests that focus on time domain analysis, we investigate structural stability in frequency domain. The intuition is straightforward: if structural change exists, then estimation methods based on the whole sample will miss it because the OLS or 2SLS estimator cannot capture the local behavior of the parameters at each time point. Consequently, the estimated residuals will contain such information. By projecting the estimated residuals on the frequency domain using the Discrete Fourier Transform (DFT), we can infer the existence of structural change by examining the DFT at each frequency. The merits of our frequency domain based approach are as follows. First, our test is consistent against both abrupt structural breaks and smooth structural changes because the Fourier transform can capture all the time series property of the unknown parameter. As long as it varies with time, our test can detect it. Second, compared to the consistent tests for smooth structural changes (e.g., Chen and Hong, 2012; Zhang and Wu, 2012; Cai et al., 2015; Chen 2015), our test avoids smoothed nonparametric estimation of the unknown parameter. As a result, our test is tuning parameter free and can detect a class of local alternatives at a faster rate. Third, we allow for endogenous covariates, and our test can be viewed as a unified framework. Moreover, our test is robust to possible structural changes in both the covariates and instruments. This is an improvement over Hall et al. (2012) and Perron and Yamamoto (2014), where the instability in the first stage reduced form has a nontrivial impact on testing and estimation. In additions, compared to the existing tests for abrupt structural breaks, e.g., Andrews' (1993) supremum test and Bai and Perron's (1998) double maximum test, we do not require trimming of the boundary region. Furthermore, unlike the existing literature that cannot distinguish structural change from model misspecification, our novel testing via the DFT is robust to model misspecification in certain cases. Therefore, the only source of rejection for our test will be the structural change. At last, our test can be extended to testing for structural change in other settings as a unified approach.

1.3 Testing Constancy of Conditional Joint Dependence

Studying the joint dependence structure has drawn great attention in both economics and statistics literature. And it is the key to modeling and testing the relationship between random variables that we are interested in. In specification testing, there is a vast number of papers on testing unconditional independence and conditional independence. While little attention has been paid to tests on the constancy of conditional joint dependence.

Suppose we have three random variables $X \in \mathbb{R}$, $Y \in \mathbb{R}$, and $Z \in \mathbb{R}$, and let $\sigma_{yz}(u,v) \equiv Cov(e^{iu'Y}, e^{iv'Z})$ be the generalized unconditional covariance function and $\sigma_{yz}(u,v,x) \equiv Cov(e^{iu'Y}, e^{iv'Z}|X_t = x)$ be the generalized conditional covariance function, where $u \in \mathbb{R}$ and $v \in \mathbb{R}$. Then we say that the conditional join dependence of Y and Z is constant with respect to X if $\sigma_{yz}(u,v) = \sigma_{yz}(u,v,x)$ for all x that belongs to the support of X_t. In this framework, the generalized unconditional covariance function $\sigma_{yz}(u,v)$ serves as a proxy for the dependence strength between Y and Z and the generalized conditional covariance function $\sigma_{yz}(u,v,x)$ measures the dependence strength between Y and Z as a function of X. If the joint dependence between Y and Z does not vary with X, then the generalized conditional covariance function will be a constant and it is equal to the generalized unconditional covariance function.

There are three reasons why we use the generalized (un)conditional covariance function as a proxy for (un)conditional joint dependence. First, the generalized (un)conditional covariance function has been intensively used in testing (un)conditional independence. For example, Székely et al. (2007) proposes a test for independence by checking if the generalized unconditional covariance function is equal to 0. Wang and Hong (2012) tests for conditional independence by checking if the generalized unconditional covariance function is equal to 0. The magnitude of the generalized (un)conditional covariance can be seen as measuring the dependence strength between Y and Z. Second, the generalized (un)conditional covariance function is based on characteristic functions and it has moment generation property (if the corresponding moment exists). If we assume $\sigma_{yz}(u,v,x)$ and $\sigma_{yz}(u,v)$ are differentiable at the origin with respect to u and v, we can take derivatives with respect to both u and v and let $u = v = 0$. Then we can compare the constancy of co-movement of at specific moments. This nice property makes our test to be able to serve the specific natures of the testing problem. It can also help us to gauge the source of rejection and provide us important informations on modeling economic relationships and making relevant inferences (Hong, 1999). Third, It is obvious that both the generalized unconditional covariance function and generalized conditional covariance function are covariance-based measures. The most common covariance-based measure is the covariance between two random variables at the first moment. One example is the regression coefficient in classical linear regression models that characterizes the co-movement between regressors and regressand. However, covariance at the first moment captures the whole dependence only when the normality condition holds and numerous studies have shown that economic variables usually do not follow a normal distribution. Using generalized (un)conditional covariance function can overcome this problem and it can capture the (un)conditional dependence at all moments if they exist.

Testing constancy of conditional joint dependence is closely connected to the study on financial contagion which is growing very fast recently. The key interest in financial contagion is to identify the factors that drive the co-movement between financial markets. Such factors are important because they may serve as the linkage between financial market through which the shocks are transmitted from one market to the other. Identifying such factors can help us to better understand the mechanism of financial contagion and make corresponding policies to prevent financial cascades. Numerous studies have been done on this topic. For example, among many, Grubel and Fadner (1971), Becker, Finnerty and Gupta (1990), Theodossiou and Lee (1993), Lin, Engle, and Ito (1994), Craig, Dravid, and Richardson (1995), Longin and Solnik (1995), Karolyi and Stulz (1996), Forbes and Rigobon (2002), etc. However the aforementioned studies only focus on the first two moments, i.e., mean and variance, and they fail detect the dependence of the whole distribution when the co-movement dynamic is beyond the first two moments. For instance, it has been shown that the dependence between stock market returns has asymmetric effects, e.g., Ang and Chen (2002), Ang and Bekaert (2000), Longin and Solnik (2001), Patton (2006), Hong, Tu, and Zhou (2007).

Detecting the factors that drive the co-movements between financial markets is equivalent to testing if the conditional joint dependence varies with the factor that we are interested in. Our test can serve this purpose by detecting the constancy of co-movement at any moment as well as the tail because the generalized (un)conditional covariance function captures the (un)conditional dependence of the whole distribution. Furthermore, as mentioned above, due to the moment generating property of the characteristic function, our test statistic can be transformed to make specific inference on what part of the joint dependence varies with the factor that we are interested in.

By designing such a test, the contribution of this essay is as the following. First, this test supplement the demand in financial contagion study in that our test can detect the factor that may serve as a link between financial markets. This test statistic capture the co-movement of the whole distribution because the characteristic function is a Fourier transform of the density function and they are equivalent to each other. Furthermore, the test statistic can be transformed to diagonalize the specific pair of moments that are driven by the factor we are interested in.

Second, the breakthrough of this test is that we can compare two groups of random variables rather than two random variables. For example, if we want to examine the dependence between Y and Z using conditional correlation, then Y or Z has to be random scalar. Otherwise the conventional correlation is not well defined. However, using the conditional characteristic function, both Y and Z can be of arbitrary dimensions and only the dimension of the conditioning factor X matters. In financial contagion study, this allows us to investigate the joint dependence between two blocks of financial markets.

Third, we show that the computationally the dimension of Y and Z doesn't matter by using a special weighting function proposed by Székely et al. (2007). The special weighting function can transform the numerical integration into L_2-norms which greatly relive the computation burden when the dimension of Y and Z are large.

Last by not least, we show that the application of the proposed test statistic is not confined to the study of financial contagion. It can also be used, with proper transformation, to test functional coefficient models (Cai et al., 2000), testing smooth structural change (Chen and Hong, 2012) and testing constant conditional correlation (Engle, 2002).

CHAPTER 2

A MODEL-FREE CONSISTENT TEST FOR STRUCTURAL CHANGE IN REGRESSION POSSIBLY WITH ENDOGENEITY

2.1 Hypotheses of Interest and Approach

Throughout this chapter, we consider the following time-varying regression model

$$Y_t = g_t(X_t) + v_t,\ t = 1, 2, ..., T, \tag{2.1}$$

where Y_t is the dependent variable, X_t is a d-dimensional vector of covariates, v_t is an unobservable disturbance, $g_t(\cdot)$ is an unknown possibly time-varying function, and T is the sample size. We allow covariates X_t to be either exogenous or endogenous. When X_t is exogenous, we have $E(v_t|X_t) = 0$. If X_t is endogenous, then $E(v_t|X_t) \neq 0$. In this case, we assume that there exists a q-dimensional vector Z_t of instrumental variables, where $q \geq d$, such that $E(v_t|Z_t) = 0$, and X_t and Z_t are not independent. We allow both covariates and instrumental variables to be either continuous or discrete. We note that estimation of a nonparametric regression model with endogeneity has attracted increasing attention in the literature (e.g. Horowitz, 2011).

Our interest is in testing the null hypothesis of no structural change in $g_t(\cdot)$:

$$\mathbb{H}_0 : g_t(\cdot) = g_0(\cdot),\ \text{for all } t = 1, 2, ..., T,$$

for some unknown time-invariant function $g_0(\cdot)$. The alternative hypothesis is that

$$\mathbb{H}_A : g_t(\cdot) \neq g_0(\cdot),\ \text{for some } t = 1, 2, ..., T,$$

for any unknown time-invariant function $g_0(\cdot)$.

Unlike the existing approaches to estimating and testing $g_t(X_t)$ by a nonparametric method, which suffers from the notorious "curse of dimensionality" problem, we propose a novel approach based on the Fourier transform. Assuming that $\{X_t, Z_t\}_{t=1}^T$ are jointly strictly stationary, we consider the Fourier transform of Eq.(2.1), namely:

$$E(Y_t e^{iu'Z_t}) = E\left[g_t(X_t)\, e^{iu'Z_t}\right] + E(v_t e^{iu'Z_t}). \tag{2.2}$$

Given $E(v_t|Z_t) = 0$, we have $E(v_t e^{iu'Z_t}) = Cov(v_t, e^{iu'Z_t}) = 0$ for all $u \in \mathbb{R}^q$. Therefore, Eq.(2.2) can be rewritten as

$$\phi_t(u) \equiv E(Y_t e^{iu'Z_t}) = E\left[g_t(X_t)\, e^{iu'Z_t}\right]. \tag{2.3}$$

The complex-valued function $\phi_t(u)$ can be viewed as the Fourier transform of the regression function $g_t(X_t)$. It is time-invariant for any $u \in \mathbb{R}^q$ if and only if $\mathbb{H}_0$ holds, as is stated in the lemma below.

Lemma 2.1.1. *$Pr[g_t(X_t) = g_0(X_t)] = 1$ for some time-invariant function $g_0(\cdot)$ if and only if $\phi_t(u) = \phi_0(u)$ for all $u \in \mathbb{R}^q$ and for some time-invariant function $\phi_0(\cdot)$.*

Intuitively, Lemma 2.1.1 says that $g_t(\cdot)$ is time-invariant if and only if its Fourier transform $\phi_t(u)$ is time-invariant. Therefore, we can rewrite the hypotheses of interest $\mathbb{H}_0$ versus $\mathbb{H}_A$ as follows:

$$\mathbb{H}_0: \qquad \phi_t(u) = \phi_0(u) \text{ for all } u \in \mathbb{R}^q, \quad \text{for all } t = 1, 2, ..., T,$$

where $\phi_0(u)$ is some unknown time-invariant function. The alternative hypothesis is

$$\mathbb{H}_A: \qquad \phi_t(u) \neq \phi_0(u) \text{ for some } t = 1, 2, ..., T, \text{ all } u \in \mathbb{R}^q \text{ in a nontrivial set,}$$

and any time-invariant function $\phi_0(u)$.

Intuitively, given that both X_t and Z_t are strictly stationary, the only source of instability of $\phi_t(u)$ is the instability of the unknown functional form $g_t(\cdot)$. Thus, we can test instability of $g_t(\cdot)$ by checking the time-varying property of its Fourier transform $\phi_t(u)$. We now consider three examples to illustrate our idea.

Example 2.1.1. [Stable Time Series Model]: *Let* $X_t = 0.5X_{t-1} + w_t$*,* $w_t \sim i.i.d.N(0, 0.75)$*, and*

$$Y_t = 1.5 + 0.5X_t + v_t,$$

where $v_t \sim i.i.d.N(0, 1)$ and $E(v_t|X_t) = 0$, which implies that X_t is exogenous, and so we have $Z_t = X_t$. Also, there exists no structural change in $g_t(X_t) = g_0(X_t) = 1.5 + 0.5X_t$. In this case, the Fourier transform

$$\phi_t(u) = [1.5 + 0.5(1 + iu)]e^{iu-\frac{u^2}{2}}$$

is time-invariant. Figures A.1 and A.2 plot the real and imaginary parts of $\phi_t(u)$ over frequency $u \in \mathbb{R}$ and time $t/T \in [0, 1]$. The shape of $\phi_t(u)$ is time-invariant.

Example 2.1.2. [Abrupt Structural Break]: *Let* $X_t = 0.5X_{t-1} + w_t$*,* $w_t \sim i.i.d.N(0, 0.75)$*, and*

$$Y_t = \begin{cases} 1.5 + 0.5X_t + v_t, & \textit{if } t \leq 0.3T, \\ -0.3X_t + v_t, & \textit{otherwise,} \end{cases}$$

where $v_t \sim i.i.d.N(0, 1)$ and $E(v_t|X_t) = 0$. The regression function is time-varying, namely

$$g_t(X_t) = \begin{cases} 1.5 + 0.5X_t, & \text{if } t \leq 0.3T, \\ -0.3X_t, & \text{otherwise.} \end{cases}$$

Its Fourier transform is given by

$$\phi_t(u) = \begin{cases} [1.5 + 0.5(1 + iu)]e^{iu - \frac{u^2}{2}}, & \text{if } t \leq 0.3T, \\ -0.3(1 + iu)e^{iu - \frac{u^2}{2}}, & \text{otherwise,} \end{cases}$$

for all $u \in \mathbb{R}$. Figures A.3 and A.4 depict the shape of $\phi_t(u)$ over $u \in \mathbb{R}$ and $t/T \in [0, 1]$, which has a level shift at $t/T = 0.3$ for its real and imaginary parts respectively. We note that the breakpoints are the same for both the real and imaginary parts.

Example 2.1.3. [Smooth Structural Change]: *Let* $X_t = 0.5X_{t-1} + w_t$, $w_t \sim i.i.d.N(0, 0.75)$, *and*

$$Y_t = \beta_0(t/T) + \beta_1(t/T)X_t + v_t,$$

with $v_t \sim i.i.d.N(0, 1)$, *where*

$$\begin{aligned} \beta_0(\tau) &= 0.2exp(-0.7 + 3.5\tau), \\ \beta_1(\tau) &= 2\tau + exp[-16(\tau - 0.5)^2] - 1. \end{aligned}$$

Now the regression function $g_t(X_t)$ changes smoothly over time:

$$g_t(X_t) = 0.2exp(-0.7 + 3.5t/T) + (2t/T + exp[-16(t/T - 0.5)^2] - 1)X_t.$$

Its Fourier transform is

$$\phi_t(u) = [0.2exp(-0.7 + 3.5t/T) + (2t/T + exp[-16(t/T - 0.5)^2] - 1)(1 + iu)]exp\left(iu - u^2/2\right)$$

for all $u \in \mathbb{R}$. In this case, $\phi_t(u)$ is a smooth function of time as is shown in Figures A.5 and A.6.

To test $\mathbb{H}_0$ against $\mathbb{H}_A$, we can estimate $\phi_t(u)$ and $\phi_0(u)$ consistently by $\hat{\phi}_t(u)$ and $\hat{\phi}_0(u)$ say, and compare their differences for all $u \in \mathbb{R}^q$. Under $\mathbb{H}_0$, the difference between $\hat{\phi}_t(u)$ and $\hat{\phi}_0(u)$ will be close to 0 for all $u \in \mathbb{R}^q$ as the sample size $T \to \infty$, and they converge to different limits under $\mathbb{H}_A$. This is similar to Hausman's (1978) test, which compares two finite dimensional parameter estimators that converge to the same limit under correct model specification and generally to different limits under model misspecification. In the present context, the main advantage of using the Fourier transform is that we avoid nonparametric estimation of unknown regression function $g_t(X_t)$. Nonparametric estimation of $\phi_t(u)$ involves only smoothing over univariate time index. In contrast, nonparametric estimation of $g_t(X_t)$ involves $(d+1)$-dimensional smoothing or even higher if the dimension of the instrument vector Z_t is higher than d. The convergence rate of such a nonparametric estimator is rather slow, especially when the dimension of X_t or Z_t is high. This would have an adverse impact on the performance of related tests, at least in finite samples. Indeed, our test is asymptotically more powerful than Vogt's (2015) consistent test under a class of local alternatives; see Section 2.4 for more discussion. To ensure the consistency of our test against $\mathbb{H}_A$, we need to check $\mathbb{H}_0$ for all $u \in \mathbb{R}^q$. This will involve integration over all $u \in \mathbb{R}^q$. The dimension of u is the same as the dimension of Z_t. However, we will show that by using a proper weighting function of u, one can avoid computationally expensive numerical integration. This makes the computation of our test statistic rather simple in practice.

A special case of our framework is when all regressors are exogenous, namely when $E(v_t|X_t) = 0$. Existing nonparametric tests for $\mathbb{H}_0$ in the literature all consider this special case. In our framework we can simply replace Z_t by X_t because $E(v_t|X_t) = 0$ implies $Cov(v_t, e^{iu'X_t}) = 0$ for all $u \in \mathbb{R}^d$. The property of our test statistic and all asymptotic results will remain the same as in the general case with endogeneity. In this chapter, our discussion is based on the general framework with endogeneity.

2.2 Nonparametric Testing

2.2.1 Nonparametric Estimation

Given the sample $\{Y_t, X_t, Z_t\}_{t=1}^T$, we first construct consistent estimators for $\phi_t(u)$ and $\phi_0(u)$. Under $\mathbb{H}_0$, $\phi_t(u) = \phi_0(u)$ is a constant function over time. We can estimate $\phi_0(u)$ consistently by its sample analog

$$\hat{\phi}_0(u) = \frac{1}{T}\sum_{t=1}^{T} Y_t e^{iu'Z_t}.$$

Throughout this chapter, we assume

$$g_t(X_t) = g\left(X_t, \frac{t}{T}\right),$$

for some function $g : \mathbb{R}^d \times [0,1] \to \mathbb{R}$. Here we assume $g_t(X_t)$ is a function of normalized time ratio $\frac{t}{T}$, rather than t directly. This implies $\phi_t(u) = \phi(u, \frac{t}{T})$ for some complex-valued function $\phi(\cdot,\cdot)$. The reason that we rescale time t to a finer grid on $[0,1]$ is to ensure that the amount of data increases around each time point t as the sample size $T \to \infty$. This is necessary for consistency of $\hat{\phi}_t(u)$. Under $\mathbb{H}_A$, we assume that for each u, $\phi(u, \frac{t}{T})$ is a twice continuously differentiable function of $\frac{t}{T}$ expect for a set of finite number of points on $[0,1]$. This set of points represents abrupt structural breaks in the unknown function $g(X_t, \frac{t}{T})$, with no prior knowledge about break dates. Our test can thus detect instability caused by both abrupt breaks as well as smooth changes; see Chen and Hong (2012) for detailed discussion.

We borrow the notion of local stationarity to characterize the instability of the predictive relationship between X_t and Y_t caused by smooth structural change. Consider the following pseudo regression:

$$Y_t e^{iu'Z_t} = \phi_t(u) + \varepsilon_t(u), \tag{2.4}$$

where $\phi_t(u) = E(Y_t e^{iu'Z_t})$ is the local mean of dependent variable $Y_t e^{iu'Z_t}$, and $\varepsilon_t(u) = Y_t e^{iu'Z_t} - \phi_t(u)$ is a generalized disturbance with $E[\varepsilon_t(u)] = 0$ for all $u \in \mathbb{R}^q$. Thus $\phi_t(u)$ can be estimated consistently by the Nadaraya-Watson (NW) estimator or the local linear (LL) estimator.

The LL estimator has an advantage over the NW estimator in the boundary region under $\mathbb{H}_A$, because the bias of local linear smoothing in the boundary region is of the same order of magnitude as that in the interior region. However, the latter is better than the former under $\mathbb{H}_0$ because $\phi_t(u)$ is constant function of $\frac{t}{T}$ and the bias for nonparametric estimators is zero for both the interior and boundary regions. In that case, the NW estimator is better since it has only one parameter to estimate.

Although the LL estimator has a better convergence rate in the boundary region over the NW estimator, the scale in the boundary region is different from that in the interior region. Let $K_h(z) = h^{-1}K(\frac{z}{h})$, where $K : [-1, 1] \to \mathbb{R}^+$ is a kernel function and $h \equiv h(T)$ is a bandwidth such that $h \to 0$ and $Th \to \infty$ as $T \to \infty$. As shown in Cai (2007), the asymptotic bias of the LL estimator in the boundary region is proportional to $h^2 b(c)$ where $b(c) = \frac{\mu_{2c}^2 - \mu_{1c}\mu_{2c}}{\mu_{0c}\mu_{2c} - \mu_{1c}^2}$, $\mu_{jc} = \int_{-c}^{-1} \eta^j K(\eta) d\eta$, for $j = 1, 2, 3$, and $c \in [0, 1]$, rather than $h^2 \int_{-1}^{1} \eta^2 K(\eta) d\eta$ in the interior region. Moreover, the asymptotic variance in the boundary region is larger than that in the interior region, which would further complicate the form of our test statistic. As discussed in Chen and Hong (2012), the boundary regions $[1, Th] \cup [T - Th, T]$ contain abundant information in finite samples. Thus, we follow Chen and Hong (2012) to use the reflection method which is introduced by Hall and Wehrly (1991) to construct pseudo data $(Y_t, X_t) = (Y_{-t}, X_{-t})$ for $t \in [1 - \lfloor Th \rfloor, 0]$ and $(Y_t, X_t) = (Y_{2T-t}, X_{2T-t})$ for $t \in [T + 1, T + \lfloor Th \rfloor]$, where $\lfloor Th \rfloor$ denotes the integer part of Th. We then use the union of the original data and the pseudo data to construct our test statistic. This method ensures that the bias in the boundary region is of the same form as that in the interior region, because we have symmetric coverage of observations in the boundary regions when the augmented data is used.

Specifically, the NW estimator for $\phi_t(u)$ with the reflection method is defined as

$$\hat{\phi}_t^{NW}(u) = \sum_{s=t-\lfloor Th \rfloor}^{t+\lfloor Th \rfloor} Y_s e^{iu'Z_s} H_{st}^{NW},$$

where $H_{st}^{NW} = \frac{K_h(\frac{s-t}{T})}{\sum_{s=t-\lfloor Th \rfloor}^{t+\lfloor Th \rfloor} K_h(\frac{s-t}{T})}$.

Alternatively, we can estimate $\phi_t(u)$ using the LL estimator $\hat{\phi}_t^{LL}(u)$. Let $\beta(u,t) = [\beta_1(u,t), \beta_2(u,t)]'$ be a 2×1 complex-valued vector such that

$$\begin{aligned} \beta_1(u,t) &= \phi_t(u), \\ \beta_2(u,t) &= \frac{\partial \phi_t(u)}{\partial(\frac{t}{T})}. \end{aligned}$$

By the Taylor expansion, for any fixed point $\frac{t}{T} \in [0,1]$, we have

$$\phi_s(u) = \beta_1(u,t) + \beta_2(u,t)\left(\frac{s-t}{T}\right) + O\left[\left(\frac{s-t}{T}\right)^2\right].$$

If $\hat{\beta}(u,t)$ is the LL estimator, then

$$\begin{bmatrix} \hat{\beta}_1(u,t) \\ \hat{\beta}_2(u,t) \end{bmatrix} = \underset{\beta(t)\in\mathbb{C}^2}{\operatorname{argmin}} \sum_{s=t-\lfloor Th \rfloor}^{t+\lfloor Th \rfloor} \left| Y_s e^{iu'Z_s} - \beta_1(u,t) - \beta_2(u,t)\left(\frac{s-t}{T}\right)\right|^2 K_h\left(\frac{s-t}{T}\right). \tag{2.5}$$

It is easy to show that the solution to Eq.(2.5) is given by

$$\begin{bmatrix} \hat{\beta}_1(u,t) \\ \hat{\beta}_2(u,t) \end{bmatrix} = \begin{bmatrix} S_{T,0}(u,t) & S_{T,1}(u,t) \\ S_{T,1}(u,t) & S_{T,2}(u,t) \end{bmatrix}^{-1} \begin{bmatrix} \Gamma_{T,0}(u,t) \\ \Gamma_{T,1}(u,t) \end{bmatrix},$$

where $S_{T,j}(u,t) = \sum_{s=t-\lfloor Th \rfloor}^{t+\lfloor Th \rfloor} \left(\frac{s-t}{T}\right)^j K_h(\frac{s-t}{T})$, $\Gamma_{T,j}(u,t) = \sum_{s=t-\lfloor Th \rfloor}^{t+\lfloor Th \rfloor} \left(\frac{s-t}{T}\right)^j K_h(\frac{s-t}{T}) Y_s e^{iu'Z_s}$. The LL estimator $\hat{\phi}_t^{LL}(u)$ is equal to the intercept estimator $\hat{\beta}_1(u,t)$ and it has an analytical solution

$$\hat{\phi}_t^{LL}(u) = \sum_{s=t-\lfloor Th \rfloor}^{t+\lfloor Th \rfloor} Y_s e^{iu'Z_s} H_{st}^{LL},$$

where $H_{st}^{LL} = \frac{[S_{T,2}(u,t) - S_{T,1}(u,t)(\frac{s-t}{T})] K_h(\frac{s-t}{T})}{S_{T,0}(u,t) S_{T,2}(u,t) - S_{T,1}(u,t)^2}$.

Both the NW and LL estimators are simply a weighted sum of $Y_s e^{iu'Z_s}$ with different weights. They are essentially asymptotically equivalent to each other since

$$\hat{\phi}_t^{LL}(u) = \frac{1}{T} \sum_{s=t-\lfloor Th \rfloor}^{t+\lfloor Th \rfloor} Y_s e^{iu'Z_s} K_h\left(\frac{s-t}{T}\right)[1 + op(1)],$$

$$\hat{\phi}_t^{NW}(u) = \frac{1}{T} \sum_{s=t-\lfloor Th \rfloor}^{t+\lfloor Th \rfloor} Y_s e^{iu'Z_s} K_h\left(\frac{s-t}{T}\right)[1 + op(1)].$$

Therefore, for notational simplicity we suppress the superscript indicating the type of weighting and write it as follows:

$$\hat{\phi}_t(u) = \sum_{s=t-\lfloor Th \rfloor}^{t+\lfloor Th \rfloor} Y_s e^{iu'Z_s} H_{st}.$$

2.2.2 Test Statistic

We propose a test statistic based on the quadratic distance between $\hat{\phi}_t(u)$ and $\hat{\phi}_0(u)$:

$$\hat{Q} = \frac{1}{T} \sum_{t=1}^{T} \int_{\mathbb{R}^q} \left| \hat{\phi}_t(u) - \hat{\phi}_0(u) \right|^2 W(u) du, \tag{2.6}$$

where $W(\cdot) : \mathbb{R}^q \to \mathbb{R}^+$ is a nonnegative, continuous, and symmetric weighting function for u at the practitioner's discretion. Unlike Su and Xiao (2008), the choice of the weighting function does not have any impact on the asymptotic distribution and consistency of our test as long as it has unbounded support.

However, different weighting functions affect the computation of the test statistic since integration over $\mathbb{R}^q$ is involved, which would be computationally demanding when the dimension of Z_t is high. Fortunately, some proper choices of $W(\cdot)$ can avoid numerical integration. One example is the joint normal density function; see Section 2.5 for more discussion. Under $\mathbb{H}_0$, the test statistic $\hat{Q}$ should converge to 0. Any significant departure of $\hat{Q}$ from 0 indicates that $g_t(\cdot)$ is time-varying.

To derive the asymptotic distribution of our test statistic, we consider a standardized version

$$\widehat{SQ} = \left(Th^{1/2}\hat{Q} - \hat{B}\right) / \sqrt{\hat{V}},$$

where

$$\hat{B} = h^{-1/2} \int_{\mathbb{R}^q} |\hat{\Omega}(u,u)| W(u) du \int K^2(\eta) d\eta$$

and

$$\hat{V} = 2 \int_{\mathbb{R}^{2q}} |\hat{\Omega}(u,v)|^2 W(u) W(v) du dv \int \left[\int K(\eta) K(\eta + \lambda) d\eta \right]^2 d\lambda$$

are consistent estimators for the asymptotic mean and asymptotic variance of $Th^{1/2}\hat{Q}$. Here $\hat{\Omega}(u,v)$ is a consistent estimator for the generalized long-run variance

$$\Omega(u,v) = \sum_{j=-\infty}^{\infty} E[\varepsilon_t(u) \varepsilon_{t-j}(v)^*].$$

We need to introduce the generalized long-run variance in both the asymptotic mean and asymptotic variance because the generalized disturbance $\varepsilon_t(u)$ is generally neither i.i.d. nor martingale difference sequence.

Following Newey and West (1987) and Andrews (1991), we can use the following consistent estimator for $\Omega(u,v)$:

$$\hat{\Omega}(u,v) = \sum_{j=-p_T}^{p_T} k_l\left(\frac{j}{p_T}\right) \hat{\sigma}_j(u,v),$$

where the jth order sample generalized covariance function is

$$\hat{\sigma}_j(u,v) = \begin{cases} \frac{1}{T}\sum_{t=j+1}^{T} \hat{\varepsilon}_t(u)\hat{\varepsilon}_{t-j}(v)^*, & j = 0, 1, ..., p_T, \\ \frac{1}{T}\sum_{t=1-j}^{T} \hat{\varepsilon}_{t+j}(u)\hat{\varepsilon}_t(v)^*, & j = -1, -2, ..., -p_T, \end{cases}$$

and the truncation lag order $p_T = p(T)$ satisfies $p_T h^{1/2} \to \infty$, $p_T/(Th) \to 0$ as $T \to \infty$. $k(\cdot) : [-1,1] \to \mathbb{R}^+$ is a symmetric, bounded, and square-integrable kernel for lag order with $k(0) = 1$.

2.3 Asymptotic Distribution

To derive the asymptotic distribution of $\widehat{SQ}$ under $\mathbb{H}_0$, we impose the following regularity conditions.

Assumption 2.3.1. *(i)$\{X_t', Z_t'\}$ is a $(d+q)\times 1$ strictly stationary β-mixing process with mixing coefficient $\{\beta(j)\}$ satisfies $\sum_{j=1}^{\infty} j^2\beta(j)^{\delta/(1+\delta)} < C$ for some $0 < \delta < 1$; and (ii) $E(Y_t^4) \leq C < \infty$.*

Assumption 2.3.2. *The instrument vector Z_t satisfies $E(v_t|Z_t) = 0$ almost surely and $F_{XZ}(x,z) \neq F_X(x)F_Z(z)$ where $F_{XZ}(\cdot)$, $F_X(\cdot)$ and $F_Z(\cdot)$ denote the joint CDF for X_t and Z_t, and the marginal CDFs for X_t and Z_t respectively.*

Assumption 2.3.3. *$K : [-1,1] \to \mathbb{R}^+$ is symmetric, bounded and twice continuously differentiable with $\int_{-1}^{1} K(u)du = 1$, $\int_{-1}^{1} uK(u)du = 0$, and $\int_{-1}^{1} u^2K(u)du = C_k < \infty$.*

Assumption 2.3.4. *$W(\cdot) : \mathbb{R}^q \to \mathbb{R}^+$ is a nonnegative, continuous, symmetric, and integrable function with $\int_{\mathbb{R}^q} |u|_q^4 W(u)du < \infty$.*

Assumption 2.3.5. *$k : [-1,1] \to \mathbb{R}^+$ is symmetric, bounded, and square-integrable with $k(0) = 1$.*

Assumption 2.3.6. *(i) The bandwidth $h = c_h T^{-\tau_h}$ for $0 < \tau_h < 1$ and $0 < c_h < \infty$; and (ii) the truncation lag order $p_T = c_p T^{\tau_p}$ for $\frac{\tau_h}{2} < \tau_p < 1 - \tau_h$ and $0 < c_p < \infty$.*

Assumption 2.3.1(i) restricts $\{X_t', Z_t'\}$ to be jointly strictly stationary. This simplifies our asymptotic analysis. It is possible to relax this assumption to allow $\{X_t', Z_t'\}$ to be locally stationary. This would allow for the locally stationary autoregressive model in Dahlhaus (1996) as well as other locally stationary nonlinear time series models. However, it is beyond the scope of this essay, and particularly our main idea here is to focus on a novel testing method based on the Fourier transform of data. The β-mixing condition is common in time series analysis and is adopted in (e.g.) Hjellvik et al. (1998), Chen and Hong (2012), Wang and Hong (2012). Assumption 2.3.1(ii) is a moment condition on Y_t, implying that the unknown function $g_t(\cdot)$ is square-integrable, which guarantees the existence of its Fourier transform.

In Assumption 2.3.2, $E(v_t|Z_t) = 0$ characterizes the validity of instruments. Also, the mutual dependence between X_t and Z_t ensures the relevance of our instruments. Otherwise, our test will have little power. Assumption 2.3.2 is necessary to ensure that our nonparametric estimator for the Fourier transform $\phi_t(u)$ is consistent at all points $\frac{t}{T}$ except for a set of a finite number of points.

Assumption 2.3.3 provides conditions on kernel $K(\cdot)$ for estimation of $\phi(u, \frac{t}{T})$. It includes but does not restrict to the Epanechnikov and Quartic kernels. Assumption 2.3.4 provides regularity conditions on weighting function $W(u)$. The choice of $W(u)$ does not affect the asymptotic distribution of the test statistic, and does not affect the consistency of the test provided $W(u)$ is a symmetric density function with unbounded support. However, the choice of $W(u)$ affects the computation of the test statistic $\widehat{SQ}$, which can be quite involved when the dimension of instruments is high. Some suitable choices of $W(\cdot)$ will avoid numerical integration, which is rather attractive in practice, especially when the dimension of Z_t is high. We will further discuss this in Section 2.5.

Assumption 2.3.5 provides conditions on kernel $k(\cdot)$ for estimation of a generalized long-run variance-covariance. Such kernels as the Bartlett and Quadratic-Spectral kernels satisfy this condition. Assumption 2.3.6 provides conditions on smoothing parameters. It implies that $h \to 0$, $p_T h^{1/2} \to \infty$, $p_T/(Th) \to 0$ as $T \to \infty$. A data-dependent method to choose an optimal h has been discussed in Robinson (1989). However, such method may affect the size of the test in finite simples because more noise is introduced. As suggested by Chen and Hong (2012), we can use a simple rule-of-thumb to choose h, namely $h = (1/\sqrt{12})T^{-1/5}$, where $1/\sqrt{12}$ is the standard deviation of $U[0,1]$, because we can approximate the grid points $\{t/T : 1, 2, ..., T\}$ by $U[0,1]$ as $T \to \infty$. There are also various ways to choose the truncation lag order p_T. One choice of p_T is discussed in Lima and Xiao (2010) and Hong, Wang, and Wang (2014):

$$p_T = min\left\{\left\lfloor\left(\frac{3T}{2}\right)^{1/3}\left(\frac{2\hat{\rho}}{1-\hat{\rho}^2}\right)\right\rfloor, \left\lfloor 8\left(\frac{T}{100}\right)^{1/3}\right\rfloor\right\}, \tag{2.7}$$

where $\lfloor A \rfloor$ denotes the integer part of A, and $\hat{\rho}$ is the estimator of the first order autocorrelation of $\{Y_t, Z_t\}$.

Now we derive the asymptotic distribution of $\widehat{SQ}$ under H_0.

Theorem 2.3.1. *Suppose Assumptions 2.3.1-2.3.6 hold. Then under $\mathbb{H}_0$, $\widehat{SQ} \xrightarrow{d} N(0,1)$ as $T \to \infty$.*

The test statistic $\widehat{SQ}$ has a convenient asymptotic null $N(0,1)$ distribution, which is asymptotically pivotal. Because $\widehat{SQ}$ diverges to $+\infty$ as $T \to \infty$ under

$\mathbb{H}_A$ (see Theorem 2.3.2 below), negative values of $\widehat{SQ}$ can occur only under $\mathbb{H}_A$ for T sufficiently large. As a result, one can use one-sided $N(0,1)$ critical values. For example, the asymptotic critical valued at the 5% significance level is 1.645.

Nevertheless, the asymptotic distribution may not provide an accurate approximation in finite samples, especially when the dependence among observation is highly persistent. Suitable bootstrap procedures can be used to obtain better approximation in finite samples. In Section 2.6, we will use simulations to examine the finite sample performance of our test statistic using both asymptotic and bootstrap critical values.

2.4 Asymptotic Power

To study the asymptotic power of the $\hat{Q}$ test, we assume the following condition:

Assumption 2.4.1. $g(\cdot,\cdot) : \mathbb{R}^q \times [0,1] \to \mathbb{R}$ *is a twice continuously differentiable function of* $\frac{t}{T}$ *expect for a set of finite number of points on* $[0,1]$, *and for this set of a finite number of points* $sup_t \| lim_{\frac{t}{T}\to\tau^+} g(X_t, \frac{t}{T}) - lim_{\frac{t}{T}\to\tau^-} g(X_t, \frac{t}{T}) \| \leq C(X_t)$, *with* $E[C^4(X_t)] < \infty$.

Assumption 2.4.1 allows for both smooth structural changes and abrupt structural breaks with unknown breakpoints. For abrupt structural breaks, the sizes of breaks are bounded by a stochastic upper bound with a finite fourth moment.

Theorem 2.4.1. *Suppose Assumptions 2.3.1-2.3.6 and 2.4.1 hold. Then for any sequence of non-stochastic constants* $\{M_T = o(T\sqrt{h})\}$, $Pr(\widehat{SQ} > M_T) \to 1$ *under* $\mathbb{H}_A$ *as* $T \to \infty$.

Theorem 2.4.1 implies that $\widehat{SQ}$ is consistent against the alternative hypothesis $\mathbb{H}_A$ at any significance level. Sharing the same merits as Chen and Hong's (2012) test, our test can detect any structural changes subject to Assumption 2.4.1, without having to know any prior information about the types of structural changes and breakpoints.

To gain insights into the proposed test, we now consider a class of local smooth structural changes:

$$\mathbb{H}_{a1} : g_t(X_t) = g_0(X_t) + \kappa_T h_t(X_t),$$

where $h_t(X_t) \equiv h(X_t, \frac{t}{T})$, and $h : \mathbb{R}^d \times [0,1] \to \mathbb{R}$ is twice continuously differentiable with respect to $\frac{t}{T}$ with $h(X_t, \cdot) \neq 0$ on a set of non-zero Borel measure set of $[0,1]$. The nonstochastic factor $\kappa_T \to 0$ as $T \to \infty$ characterizes the speed at which the departure $g_t(X_t) - g_0(X_t) = \kappa_T h(X_t, \frac{t}{T})$ converges to 0 as $T \to \infty$. Under $\mathbb{H}_{a1}$, we have

$$|\phi_t(u) - \phi_0(u)|^2 = \kappa_T^2 \left| E\left[h_t(X_t) e^{iu'Z_t} \right] \right|^2.$$

Theorem 2.4.2. *Suppose Assumptions 2.3.1-2.3.6 hold. Then under $\mathbb{H}_{a1}$ with $\kappa_T = T^{-1/2}h^{-1/4}$,*

$$Pr\left[\widehat{SQ} \geq z_\alpha \middle| \mathbb{H}_{a1}\right] \to 1 - \Phi\left(z_\alpha - \frac{\gamma_1}{\sqrt{V}}\right)$$

as $T \to \infty$, where $\Phi(\cdot)$ is the $N(0,1)$ CDF, z_α is the one-sided critical value of $N(0,1)$ at significance level α,

$$V = 2\int_{\mathbb{R}^{2q}} |\Omega(u,v)|^2 W(u)W(v)dudv \int \left[\int K(\eta)K(\eta + \lambda)d\eta \right]^2 d\lambda,$$

$\Omega(u,v) = \sum_{j=-\infty}^{\infty} E[\varepsilon_t(u)\varepsilon_{t+j}(u)]$, and

$$\gamma_1 = \int_{\mathbb{R}^q} \left\{ \int_0^1 \left| E\left[h(X_t,\tau)e^{iu'Z_t} \right] \right|^2 d\tau - \left| \int_0^1 E\left[h(X_t,\tau)e^{iu'Z_t} \right] d\tau \right|^2 \right\} W(u)du < \infty.$$

Theorem 2.4.2 implies that $\widehat{SQ}$ has nontrivial power under the class of local alternatives $\mathbb{H}_{a1}$ with a rate of $\kappa_T = T^{-1/2}h^{-1/4}$, which is slightly slower than the parametric rate $T^{-1/2}$. For example, if the bandwidth $h \propto T^{-1/5}$, then the rate $\kappa_T = T^{-1/2+1/20}$ is rather close to the parametric rate $T^{-1/2}$. We emphasize that the rate $\kappa_T = T^{-1/2}h^{-1/4}$ is not affected by the dimension of instruments Z_t or covariates X_t, thus free of the "curse of dimensionality" problem. In contrast, Vogt's (2015) consistent test involves smoothing of dimension $d+1$ and so suffers from the "curse of dimensionality" problem, especially when the dimension d of covariates is high. Vogt's (2015) consistent test has nontrivial asymptotic power under the class of local alternatives $\mathbb{H}_{a1}$ with a rate of $\kappa_T = T^{-1/2}h^{-(d+1)/4}$, which is slower than the rate of $\kappa_T = T^{-1/2}h^{-1/4}$, especially for a large d. Moreover, Vogt (2015) only considers testing structural changes in regression with exogenous covariates.

We note that the nonparametric tests of Hidalgo (1995) and Su and Xiao (2008) can detect the class of local alternatives $\mathbb{H}_{a1}$ with parametric rate $\kappa_T = T^{-1/2}$. However, these tests are not consistent for all departures $h_t(X_t)$. For example, Hidalgo's (1995) test has no power against $\mathbb{H}_{a1}$ when $E[\varsigma(X_t)f^2(X_t)h_t(X_t)] = 0$, where $\varsigma(X_t) = \int g(X_t, \tau)d\tau$. And Su and Xiao's (2008) tests have no power against $\mathbb{H}_{a1}$ when $E[\omega(X_t)f(X_t)h_t(X_t)] = 0$, where $f(X_t)$ is the density function of X_t and $\omega(X_t)$ is a weighting function chosen by practitioners. In contrast, our $\widehat{SQ}$ test is consistent for all departures $h_t(X_t)$.

On the other hand, suppose we consider a class of non-smooth local alternatives at some given point $\tau_0 \in [0, 1]$:

$$\mathbb{H}_{a2} : g_t(X_t) = g_0(X_t) + a_T l\left(X_t, \frac{t/T - \tau_0}{b_T}\right),$$

where τ_0 is a fixed point in $[0, 1]$, $l : \mathbb{R}^{d+1} \to \mathbb{R}$ is twice continuously differentiable with respect to τ. We let $a_T = a(T) \to 0$, and $b_T = b(T) \to 0$ as $T \to \infty$. Under $\mathbb{H}_{a2}$, $g_t(X_t)$ has a non-smooth spike at point τ_0 as $T \to \infty$. The rate $b_T \to 0$ controls the sharpness of the structural change around τ_0, and $a_T \to 0$ is the speed at which the departure of $g_t(X_t)$ from $g_0(X_t)$ at each point t/T vanishes to 0 as $T \to \infty$. This type of alternatives was first studied by Rosenblatt (1975). Chen and Hong (2012) also considers this class of local alternatives in testing parameter instability in a linear time series model.

Now we derive the asymptotic distribution of $\widehat{SQ}$ under $\mathbb{H}_{a2}$.

Theorem 2.4.3. *Suppose Assumptions 2.3.1-2.3.6 hold. Then under $\mathbb{H}_{a2}$ with $a_T \to 0$, $b_T \to 0$, $a_T^2 b_T = T^{-1}h^{-1/2}$, and $h = o(b_T)$, we have*

$$Pr\left[\widehat{SQ} \geq z_\alpha \middle| \mathbb{H}_{a2}\right] \to 1 - \Phi\left(z_\alpha - \frac{\gamma_2}{\sqrt{V}}\right)$$

as $T \to \infty$, where $\Phi(\cdot)$ is the $N(0,1)$ CDF, z_α is the one-sided critical value of $N(0,1)$ at significance level α,

$$V = 2\int_{\mathbb{R}^{2q}} |\Omega(u,v)|^2 W(u)W(v)dudv \int\left[\int K(\eta)K(\eta+\lambda)d\eta\right]^2 d\lambda,$$

$\Omega(u,v) = \sum_{j=-\infty}^{\infty} E[\varepsilon_t(u)\varepsilon_{t+j}(u)]$, *and*

$$\gamma_2 = \int_{\mathbb{R}^q}\int_{\mathbb{R}} \left|E\left[l(X_t,\eta)e^{iu'Z_t}\right]\right|^2 d\eta W(u)du < \infty.$$

Theorem 2.4.3 implies that our test has nontrivial asymptotic power under the class of non-smooth local alternatives $\mathbb{H}_{a2}$ with $a_T^2 b_T = T^{-1}h^{-1/2}$. In contrast, the nonparametric tests of Hidalgo (1995) and Su and Xiao (2008) may have no power against this class of non-smooth local alternatives, because it can be shown that these tests can detect $\mathbb{H}_{a2}$ with $a_T b_T = T^{-1/2}$ or slower. If $a_T = T^{-7/20}(\ln\ln T)^{-\epsilon/2}$, $b_T = T^{-1/5}(\ln\ln T)^{\epsilon}$, $h = T^{-1/5}$, where $\epsilon > 0$ is a constant, then we have $a_T^2 b_T = T^{-1}h^{-1/2}$ but $a_T b_T = o(T^{-1/2})$. In this case, our test has power against $\mathbb{H}_{a2}$ but the tests of Hidalgo (1995) and Su and Xiao (2008) have no power. See Chen and Hong (2012, footnote 12, P.1168) for details.

2.5 Weighting Function

We now discuss the choice of weighting function $W(\cdot)$. Theorem 2.3.1 implies that the test statistic $\widehat{SQ}$ always has an asymptotic $N(0,1)$ distribution under $\mathbb{H}_0$, no matter what weighting function $W(u)$ is used. Theorem 2.4.1 implies that our test is always consistent against $\mathbb{H}_A$ provided a continuous and integrable weighting function with unbounded support is used. However, the choice of $W(u)$ affects the computation of the test statistic $\widehat{SQ}$, which involves q-dimensional integration, where q is the dimension of instruments Z_t, which could be rather demanding when q is large.

In this section, we will discuss two types of weighting functions for $W(u)$ that avoid numerical integration, which is quite attractive in practice, especially when the dimension q is large. The first type of weighting function is the joint independent normal probability density function:

$$W_N(u) = \prod_{k=1}^{q} \frac{1}{\sqrt{2\pi\xi_k}} exp\left(-\frac{u_k^2}{2\xi_k^2}\right),$$

where ξ_k can be viewed as the standard deviation of $W_N(u)$ for dimension k. A larger ξ_k implies that higher weights are given to the values of u that are distant from 0. Hong, Wang, and Wang (2014) propose this weighting function and set $\xi_k = 1$ for all $k = 1, 2, ..., q$. Under $W_N(u)$, we have the following identity

$$\int_{\mathbb{R}} cos(u_k z_k) \frac{1}{\sqrt{2\pi\xi_k}} exp\left(-\frac{u_k^2}{2\xi_k^2}\right) du_k = exp\left(-\frac{z_k^2 \xi_k^2}{2}\right), \; k = 1, 2, ..., q. \tag{2.8}$$

For simplicity, we set $\xi_k = \xi$ for all k. Therefore, Eq.(2.6) can be written as

$$\hat{Q}_N = \frac{1}{T}\sum_{t=1}^{T}\left[\sum_{r,s\in\mathbb{S}_1} A_{sr}H_{st}H_{rt} + \frac{1}{T^2}\sum_{r,s\in\mathbb{S}_2} A_{sr} - \frac{2}{T}\sum_{s\in\mathbb{S}_1,r\in\mathbb{S}_2} A_{sr}H_{st}\right], \tag{2.9}$$

where $A_{sr} = Y_sY_re^{-\frac{|Z_s-Z_r|_q^2\xi^2}{2}}$, $|\cdot|_q$ is the Euclidean norm in $\mathbb{R}^q$, $\mathbb{S}_1 = \{-\lfloor Th\rfloor, -\lfloor Th\rfloor + 1, 2, ..., T + \lfloor Th\rfloor\}$, and $\mathbb{S}_2 = \{1, 2, ..., T\}$. Furthermore, our $\widehat{SQ}$ statistic becomes

$$\widehat{SQ} = \widehat{SQ}_N \equiv \left(Th^{1/2}\hat{Q}_N - \hat{B}_N\right)/\sqrt{\hat{V}_N} \tag{2.10}$$

where

$$\hat{B}_N = h^{-1/2}\sum_{j=-p_T}^{p_T} k\left(\frac{j}{p_T}\right)\hat{\sigma}_N(j)\int K^2(\eta)d\eta,$$

$$\hat{V}_N = 2\sum_{j,k=-p_T}^{p_T} k\left(\frac{j}{p_T}\right)k\left(\frac{l}{p_T}\right)\hat{\sigma}_N(j,l)\int\left[\int K(\eta)K(\eta+\lambda)d\eta\right]^2 d\lambda,$$

with

$$\hat{\sigma}_N(j) = \frac{1}{T}\sum_{s=1+|j|}^{T}\hat{\zeta}(s, s-j),$$

$$\hat{\sigma}_N(j,l) = \frac{1}{T^2}\sum_{s=1+|j|}^{T}\sum_{r=1+|l|}^{T}\hat{\zeta}(s,r)\hat{\zeta}(s-j, r-l).$$

Here $K_{sr} = \frac{1}{T}K_h(\frac{s-r}{T})$ and

$$\hat{\zeta}(j,k) = A_{j,k} - \frac{1}{T}\sum_{l=1}^{T}A_{jl} - \frac{1}{T}\sum_{l=1}^{T}A_{kl} + \frac{1}{T^2}\sum_{l_1,l_2}^{T}A_{l_1l_2}.$$

No numerical integration is involved in computing $\widehat{SQ}_N$, no matter how large the dimension of Z_t is.

Alternatively, we can also use another type of weighting function based on the Laplace density function:

$$W_L(u) = \prod_{k=1}^{q}\frac{\lambda_k}{2}e^{-\lambda_k|u_k|},$$

where the λ_k's are scale parameters. The Laplace density function also assigns the highest weight to $u = \mathbf{0}$ but the weights decreases rather fast as u deviates from $\mathbf{0}$. By the identity

$$\int_{\mathbb{R}} cos(u_k z_k)\frac{\lambda_k}{2}e^{-\lambda_k|u_k|} = \frac{\lambda_k^2}{\lambda_k^2 + z_k^2},\ k = 1, 2, ..., q, \tag{2.11}$$

we can rewrite Eq.(2.6)

$$\hat{Q}_L = \frac{1}{T}\sum_{t=1}^{T}\left[\sum_{r,s\in\mathbb{S}_1} A_{sr}H_{st}H_{rt} + \frac{1}{T^2}\sum_{r,s\in\mathbb{S}_2} A_{sr} - \frac{2}{T}\sum_{s\in\mathbb{S}_1,r\in\mathbb{S}_2} A_{sr}H_{st}\right], \tag{2.12}$$

where $A_{sr} = Y_s Y_r \prod_{k=1}^{q} \frac{\lambda_k^2}{\lambda_k^2+(Z_{sk}-Z_{rk})^2}$, $\mathbb{S}_1 = \{-\lfloor Th\rfloor, -\lfloor Th\rfloor + 1, 2, ..., T + \lfloor Th\rfloor\}$, and $\mathbb{S}_2 = \{1, 2, ..., T\}$. Furthermore, the $\widehat{SQ}$ statistic becomes

$$\widehat{SQ} = \widehat{SQ}_L \equiv \left(Th^{1/2}\hat{Q}_L - \hat{B}_L\right) / \sqrt{\hat{V}_L} \tag{2.13}$$

where

$$\hat{B}_L = h^{-1/2}\sum_{j=-p_T}^{p_T} k\left(\frac{j}{p_T}\right)\hat{\sigma}_L(j)\int K^2(\eta)d\eta,$$

$$\hat{V}_L = 2\sum_{j,k=-p_T}^{p_T} k\left(\frac{j}{p_T}\right)k\left(\frac{l}{p_T}\right)\hat{\sigma}_L(j,l)\int\left[\int K(\eta)K(\eta+\lambda)d\eta\right]^2 d\lambda,$$

with

$$\hat{\sigma}_L(j) = \frac{1}{T}\sum_{s=1+|j|}^{T}\hat{\zeta}(s, s-j),$$

$$\hat{\sigma}_L(j,l) = \frac{1}{T^2}\sum_{s=1+|j|}^{T}\sum_{r=1+|l|}^{T}\hat{\zeta}(s,r)\hat{\zeta}(s-j, r-l).$$

where $K_{sr} = \frac{1}{T}K_h(\frac{s-r}{T})$ and

$$\hat{\zeta}(j,k) = A_{j,k} - \frac{1}{T}\sum_{l=1}^{T}A_{jl} - \frac{1}{T}\sum_{l=1}^{T}A_{kl} + \frac{1}{T^2}\sum_{l_1,l_2}^{T}A_{l_1l_2}.$$

Again, no numerical integration is involved in computing the test statistic $\widehat{SQ}_L$.

We note that Székely et al. (2007) propose the weighting

$$W_D(u) = \frac{1}{c_m|u|_m^{m+1}},$$

where $c_m = \frac{\pi^{(1+m)/2}}{\Gamma((1+m)/2)}$ and $\Gamma(\cdot)$ denotes the Gamma function. This weighting function avoids numerical integration for their Covariance of Distance measure for independence, which is based on the empirical characteristic function. Unfortunately, the weighting function $W_D(u)$ is not integrable and does not satisfy Assumption 2.3 4. Consequently, it cannot be used in our test.

2.6 Simulation Studies

In this section, we study the finite sample performance of our test in comparison with Su and Xiao's (2008) nonparametric CUSUM-type tests. We consider the cases that covariates are exogenous and endogenous respectively.

2.6.1 The Exogeneity Case

To study the size performance, we first consider the following DGPs:

$$\begin{aligned}
\text{DGP S.1:}\ Y_t &= 1 + 1.5X_{1t} + v_t; \\
\text{DGP S.2:}\ Y_t &= 1 + 1.5X_{1t} + X_{1t}^2 + v_t \\
\text{DGP S.3:}\ Y_t &= 1 + 0.5X_{1t} - 1.5X_{2t} + X_{3t}^2 + v_t; \\
\text{DGP S.4:}\ Y_t &= 1 + 0.5X_{1t} + 2X_{4t} + v_t; \\
\text{DGP S.5:}\ Y_t &= 1 + 0.5X_{1t}X_{5t} + v_t;
\end{aligned}$$

where

$$X_{1t} = 0.5X_{1(t-1)} + v_{1t};$$

$$
\begin{aligned}
X_{2t} &= 1 - 0.5X_{2(t-1)} + v_{2t}; \\
X_{3t} &= 0.4X_{1(t-1)} + v_{3t}; \\
X_{4t} &= \begin{cases} 0 & \text{with } Pr = 0.3, \\ 1 & \text{with } Pr = 0.7; \end{cases} \\
X_{5t} &= \begin{cases} 1 & \text{with } Pr = 0.4, \\ 2 & \text{with } Pr = 0.6; \end{cases} \\
v_{2t} &\sim i.i.d.N(0,1); \\
v_{3t} &\sim i.i.d.N(0,1).
\end{aligned}
$$

DGP S.1-S.5 are stationary time series models. Under DGP S.1, $g_t(X_t)$ is a linear function. Under DGP S.2, $g_t(X_t)$ is nonlinear. Under DGP S.3, $g_t(X_t)$ is a nonlinear function with a relatively high dimension ($d = 4$). Under DGP S.4 and DGP S.5, one covariate is discrete. The discrete regressor is of an additive form under DGP S.4 and a multiplicative form under DGP S.5. We note that when there are discrete covariates, Su and Xiao's (2008) tests are no longer applicable.

To examine the power performance, we consider the following DGPs:

DGP P.1 [Single Structural Break]:

$$
Y_t = \begin{cases} 1.5 + X_{1t} + v_t & \text{if } t \le 0.3T, \\ 1.5 - 2X_{1t} + v_t & \text{otherwise,} \end{cases}
$$

DGP P.2 [Multiple Structural Breaks]:

$$
Y_t = \begin{cases} 1.5 + X_{1t} + v_t & \text{if } 0.1T < t < 0.3T, \\ 1.2 - X_{1t}^2 + 2sin(X_{1t}) + v_t & \text{if } 0.3T \le t \le 0.7T, \\ -2X_{1t} + v_t & \text{otherwise,} \end{cases}
$$

DGP P.3 [Smooth Structural Change]:

$$Y_t = \theta(\tau)(1 + 0.5X_{1t}) + v_t,$$

where $\theta(\tau) = 1.5\tau - exp(-(\tau - 0.5)^2)$ with $\tau = t/T$; DGP P.4 [Smooth Structural Change]:

$$Y_t = \theta_1(\tau)X_{1t} + exp(\tau X_{1t}) - \theta_2(\tau)X_{1t}^2 + v_t,$$

where $\tau = t/T$,

$$\begin{aligned} \theta_1(\tau) &= 0.2exp(-0.7 + 3.5\tau), \\ \theta_2(\tau) &= 2\tau + exp[-16(\tau - 0.5)^2] - 1. \end{aligned}$$

Under DGP P.1, there exists a single break. Under DGP P.2, there exist multiple breaks. Under DGP P.3 and DGP P.4, there exist smooth structural changes in linear and nonlinear regressions respectively.

For each DGP, we simulate 1000 data sets with the sample size $T = 100$ and 200 respectively. We choose the joint $N(0, 1)$ density function for $W(u)$, and the Epanechnikov kernel for $K(\cdot)$. For the choice of bandwidth, we follow Chen and Hong (2012) to use the simple rule-of-thumb bandwidth $h = (1/\sqrt{12})T^{-1/5}$. Our test statistic involves estimation of a generalized long run variance-covariance, so we need to choose a proper lag order p_T. We follow the data-dependent lag order method used by Hong, Wang, and Wang (2014) which is given by Eq.(2.7). We use the Bartlett kernel for $k(\cdot)$.

We compare our test with the nonparametric CUSUM-type tests proposed by Su and Xiao (2008). Their test statistics are are based on nonparametric estimation of the unknown regression function. We consider both types of test statistics:

$$\begin{aligned} KS_T &= \max_{1\le t\le T}\left|\frac{1}{\sqrt{T}\hat{\sigma}}\sum_{j=1}^{t}\hat{v}_j\right|, \\ CM_T &= \frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{\sqrt{T}\hat{\sigma}}\sum_{j=1}^{t}\hat{v}_j\right)^2, \end{aligned}$$

where $\hat{v}_t = [Y_t - \hat{g}(X_t)]\hat{f}(X_t)\omega(X_t)$. Here $\hat{g}(X_t)$ and $\hat{f}(X_t)$ are the nonparametric kernel estimators for the regression function $g_0(X_t)$ under $\mathbb{H}_0$ and the probability density function $f(X_t)$ of X_t respectively, and $\omega(X_t)$ is a weighting function that has substantial impact on the asymptotic power. We follow Su and Xiao (2008) to set $\omega(x) = [sin(x) + cos(x)]a_x$, where $a_x = \mathbf{1}[\hat{f}(x) > 0.001/logT]$, and choose the fourth order Epanechnikov kernel. For the bandwidth h, we choose $h = h_0 T^{\frac{1}{9}} T^{-\gamma}$ where $\gamma = \frac{1}{7}$, and h_0 is chosen via a least squares cross-validation procedure.

In addition to using asymptotic critical values, we also consider the following moving block bootstrap (MBB) method: Step (i), choose a block length $l_T = l(T)$ such that $l_T \to \infty$ as $T \to \infty$; Step (ii), divide data into $T - l_T + 1$ blocks and generate block data $\{\Xi_t\}_{t=1}^{T-l_T+1}$, where $\Xi_t = \{[Y_t, Z_t], ..., [Y_{t+l_T-1}, Z_{t+l_T-1}]\}$; Step (iii), resample $\{\Xi_t\}_{t=1}^{T-l_T+1}$ with replacement to form a bootstrap data set $\{\Xi_t^*\}_{t=1}^{L}$ satisfying $T = \lfloor Ll_T \rfloor$; Step (iv), calculate $\widehat{SQ}^*$ using $\{\Xi_t^*\}_{t=1}^{L}$; Step (v), repeat Step (iii) to (iv) B times to generate bootstrapped test statistics $\{\widehat{SQ}_b^*\}_{b=1}^{B}$. Then the p-value is given by

$$p^* = \frac{1}{B}\sum_{b=1}^{B}\mathbf{1}(\widehat{SQ} \le \widehat{SQ}_b^*).$$

We reject $\mathbb{H}_0$ when p^* is smaller than the given significance level. Choosing an optimal block length is crucial and many approaches have been proposed (e.g., Lahiri, 1999). In this essay, we adopt Politis and White's (2004) automatic block-length selection procedure.

Table A.1 reports the size performance of our test and Su and Xiao's (2006) nonparametric CUSUM-type tests. $\widehat{SQ}^{AS}$ and $\widehat{SQ}^{BS}$ denote the results of $\widehat{SQ}$ using asymptotic and bootstrap critical values respectively. For DGP S.1 and DGP S.2, we find that all tests perform reasonably well in finite samples. For DGP S.3, when the dimension of regressors is high, we find that our test displays some under rejections using asymptotic critical values, but it improves as the sample size T increases. In both DGP S.4 and DGP S.5 where one regressor is discrete, our test performs reasonably well, but Su and Xiao's (2008) test is no longer applicable.

Table A.2 reports the power performance of the tests. Not surprisingly, Su and Xiao's (2008) tests outperform our test under DGP P.1 and DGP P.2 where their tests have nontrivial power. Even so, our test is also rather powerful to detect the deviations from $\mathbb{H}_0$ under the two alternatives. Under DGP P.3 and DGP P.4, Su and Xiao's (2008) tests have little or low power. In contrast, our test has reasonable power and is more powerful than Su and Xiao's (2008) tests. It is because our test is consistent against all alternatives while Su and Xiao's (2008) tests are not consistent tests. Thus, our test is expected to have all-round power against various alternatives, although it may not have the best power against certain directions.

2.6.2 The Endogeneity Case

To study the finite sample performance of our test with endogeneity, we consider the following DGP.

DGPIV S.1 [No Structural Change]:

$$Y_t = \ln(|X_t - 1| + 1)sgn(X_t - 1) + v_t,$$

$$X_t = \beta Z_t + w_t,$$

where v_t and w_t are generated by the bivariate normal distribution

$$\begin{pmatrix} v_t \\ w_t \end{pmatrix} \sim i.i.d.N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}\right].$$

Z_t is an instrumental variable and follows an AR(1) process.

$$Z_t = 0.5Z_{t-1} + \xi_t,$$

where $\xi_t \sim$ i.i.d.$N(0, 0.75)$. This DGP is the time series version of the example used by Newey and Powell (2003), where they assume $Z_t \sim$ i.i.d.$N(0, 1)$ with $\beta = 1$. The parameter β represents the strength between the instrumental variable Z_t and the regressor X_t because their correlation is positively related to β:

$$corr(X_t, Z_t) = \sqrt{\frac{1}{1 + 1/\beta^2}}.$$

We set $\beta = 0.1$, 1, and 2 respectively to investigate the impact of weak and strong instruments on our test.

To examine the power performance, we consider the following DGPs:

DGPIV P.1 [Abrupt Structural Break]:

$$Y_t = \begin{cases} \ln(|X_t - 1| + 1)sgn(X_t - 1) + v_t & \text{if } t \leq 0.3T, \\ \ln(|X_t - 1| + 2) + v_t & \text{otherwise;} \end{cases}$$

DGPIV P.2 [Smooth Structural Change]:

$$Y_t = \theta(\tau)\ln(|X_t - 1| + 1)sgn(X_t - 1) + v_t,$$

where $\theta(\tau) = 1.5\tau - exp(-(\tau - 0.5)^2)$ with $\tau = t/T$.

Here we only examine the performance of our test because Su and Xiao's (2008) tests are not applicable to regressions with endogeneity. Tables A.3 and A.4 report the empirical rejection rates at 10% and 5% significance levels, respectively. Under DGPIV S.1, we find that when the instrumental variable has a weak dependence (e.g., $\beta = 0.1$) with the covariate, $\widehat{SQ}$ is undersized based on asymptotic critical values, but it improves as the sample size T increases. When the dependence between the instrumental variable and the covariates becomes stronger ($\beta = 2$), the size of our test becomes better. The empirical rejection rate using bootstrapped critical values closer to the nominal level than that using asymptotic critical values, especially when the sample size is small. Overall, we observe that the strength of correlation between X_t and Z_t plays a more important role than the sample size. When $\beta = 2$, the finite sample rejection rates of both $\widehat{SQ}^{AS}$ and $\widehat{SQ}^{BS}$ are close to the nominal levels. For power performance, both $\widehat{SQ}^{AS}$ and $\widehat{SQ}^{BS}$ are powerful in detecting the alternative DGPIV P.1. We observe that $\widehat{SQ}^{AS}$ is more powerful than $\widehat{SQ}^{BS}$. Our tests have low power under DGPIV P.2, but the rejection rate of $\widehat{SQ}^{BS}$ can achieve 28.4% at the 5% significance level when $\beta = 2$ and $T = 200$.

To sum up, we demonstrate good performance of our test in cases with exogeneity and endogeneity in finite samples. When covariates are all exogenous, our test is more powerful under several alternatives than Su and Xiao's (2008) tests. Moreover, when endogenous or discrete covariates are present, our test also performs reasonably well in finite samples. The existing nonparametric tests in the literature are not applicable in this case.

2.7 Empirical Application

In the section, we revisit the predictability of equity premium. It has been documented that many financial and macroeconomic variables usually have poor out-of-sample predictive power for equity returns, see (e.g.) Goyal and Welch (2008). Many studies attribute the failure of out-of-sample prediction to the existence of structural changes. Chen and Hong (2012) examine the stability of the predictive regression using 14 financial and economic variables, and find substantial evidence against the stability of both univariate and multivariate linear predictive models for equity returns. However, as Chen and Hong (2012) mention, the rejection could be due to model misspecification since they consider linear regression models. Hillebrand et al. (2009) and Campbell and Thompson (2008) find that by imposing sign restrictions, linear predictive models could provide improved out-of-sample prediction. This implies the failure of out-of-sample prediction could be a result of model misspecification. Lee, Tu, and Ullah (2014) impose monotonicity in both nonparametric and semiparametric predictive regression models, and they find that these models have better predictability for equity premium.

We now examine the following crucial question: Is the failure of out-of-sample prediction due to the existence of neglected nonlinearity or the existence of structural changes? Although Lee, Tu, and Ullah (2014) show the superior performance of nonparametric and semiparametric models over linear regression models, they might overlook the instability of the underlying DGP. We apply our test to detect structural changes that are robust to model misspecification.

We follow Lee, Tu, and Ullah (2014) and examine four predictors: default spread (ds), smoothed earning-price ratio (se/p), long-term yields on U.S. government bonds (lty), and yields on the 3-Month T-bill on the secondary market (t-bill). The dependent variable is the monthly return of the S&P 500 index (Y_{t+1}), where $Y_{t+1} = \log[(P_{t+1} + D_{t+1})/P_t] - r_t$, P_t is the monthly S&P 500 index, D_t is the dividend paid on the S&P 500 index, r_t is the 3-month Treasury bill rate. We consider the predictive regression

$$Y_{t+s} = g(X_t) + v_{t+s},$$

where s is the number of steps ahead in out-of-sample prediction. Our data are from Lee, Tu, and Ullah (2014) and Chen and Hong (2012).

To account for the oil shocks in the 1970s, we divide our sample into two subsets: the pre-oil-shock sample (January 1950 to December 1975) and the post-oil-shock sample (January 1976 to December 2005). The p-values of our test $\widehat{SQ}_N$ are obtained by the block-bootstrap as described in Section 2.6. Table A.5 reports the p-values of our test for stability of equity returns prediction. When the forecasting step is one month ahead ($s = 1$), our test rejects the stability of the predictive relationship between equity returns and each predictor, for both the pre-oil-shock and post-oil-shock periods. Although the near past predictors should contain relevant information about the outcome one step ahead, such relationship is usually unstable due to the rapidly changing economic environment. When $s = 6$, we find that the predictive relationship is unstable for the pre-oil-shock period but stable for the post-oil-shock period at the 5% significance level. When the forecasting is one year ahead, only the earning-price ratio (se/p) has an unstable predictive relationship with the future equity return at the 5% significance level in the pre-oil-shock period. The predictive relationship with each predictor is stable in the post-oil-shock period.

Different from the strong evidence of unstable predictive relationships between equity returns and financial/macroeconomic variables detected by Chen and Hong (2012), our findings indicate that the source of rejection in Chen and Hong (2012) could come from model misspecification. Moreover, our test does document strong evidence of structural changes in predictive regressions for $s = 1$ and $s = 6$ in the pre-oil-shock period. This implies that structural change exists and is one important source for poor out-of-sample forecasts, no matter what predictive regression models are used.

2.8 Conclusion

The failure of out-of-sample forecasting of stationary time series regression models may be due to the existence of neglected nonlinearity or structural changes. It is important to distinguish structural changes from neglected nonlinearity or model misspecification, because their implication on modeling, inference and prediction are different. Therefore, it is highly desirable to develop consistent tests for structural changes that are robust to model misspecification.

In this chapter, we have proposed a model-free consistent test for structural changes in regression by testing the time-varying property of the Fourier transform of data. This avoids direct nonparametric estimation of the unknown regression function, which is adopted by the existing nonparametric tests for structural changes and suffers from the notorious "curse of dimensionality" problem, especially when the dimension of regressors is high. As a result, our test is asymptotically locally more powerful than Vogt's (2015) nonparametric test for structural changes, which is the only consistent test for structural changes in a nonparametric regression model in the existing literature. Although the nonparametric tests of Hidalgo (1995) and Su and Xiao (2008) are asymptotically more powerful than our test against certain smooth local alternatives, they are not consistent tests and are asymptotically less powerful than our test against a class of non-smooth local alternatives. Unlike the existing literature, our approach is applicable to regression models with either exogenous or endogenous covariates, and we allow covariates and instruments to be discrete random variables. Because of the Fourier transform, our test statistic involves integration with the same dimension as that of the instruments, which would be rather involved when the dimension of instruments is high. By using a suitable weighting function, we can avoid numerical integration, which is computationally convenient and efficient in practice. Our test statistic follows a convenient asymptotic null $N(0, 1)$ distribution, and is consistent against a larger class of smooth structural changes as well as abrupt structural breaks with unknown breakpoints. Simulations show that the test has reasonable size and all-round power against various alternatives of abrupt breaks and smooth changes in linear and nonlinear time series regression models. An empirical application to predictive regressions of equity returns suggests that the structural changes are one important source of poor out-of-sample productivity for S&P 500 returns, no matter what parametric predictive models are employed.

CHAPTER 3

CONSISTENT TESTING FOR STRUCTURAL CHANGE IN TIME SERIES REGRESSION MODELS VIA THE FOURIER TRANSFORM

3.1 Framework and Approach

In this chapter, we consider testing for structural change of unknown form in the following linear time series regression model:

$$Y_t = X_t'\beta_t + \varepsilon_t, \tag{3.1}$$

where $\{X_t, Y_t\}_{t=1}^T$ is a $\mathbb{R}^d \times \mathbb{R}$-valued observable random sample, $\beta_t \in \Theta$ is a $\mathbb{R}^d$-valued unknown parameter vector that can vary with time, Θ denotes the parameter space which is bounded, and ε_t is an unobservable disturbance such that $E(\varepsilon_t|X_t) = 0$.

The hypothesis of interest is

$$\mathbb{H}_0 : \beta_t = \beta_0 \text{ for some unknown } \beta_0 \in \Theta \text{ and for all } t = 1, 2, ..., T,$$

against

$$\mathbb{H}_A : \beta_t \text{ is a time-varying parameter.}$$

Under $\mathbb{H}_0$, the true model parameter is equal to β_0, and there is no structural change. While under $\mathbb{H}_A$, there exist abrupt structural breaks if β_t is a step function of time; and there exist smooth structural changes if $\beta_t \equiv \beta(\frac{t}{T})$ is a smooth function of the rescaled time $\frac{t}{T} \in (0, 1]$.

A straightforward approach to testing $\mathbb{H}_0$ is to first get consistent estimates for β_t under the null and alternative hypothesis respectively. Then compare their difference over various time points, e.g., Chen and Hong (2012), Kristensen (2012), Zhang and Wu (2012), and Cai et al. (2015). These tests are generalized Hausman's test (Hausman, 1978) since the two estimators converge to the same probability limit only under $\mathbb{H}_0$. However, such testing requires choosing tuning parameters, i.e., a bandwidth, when estimating the unrestricted model. Even though the choice of a bandwidth has a trivial impact asymptotically, it may be crucial in a finite sample. It is possible that two practitioners may obtain conflicting results by using two different bandwidths. To circumvent the undesirable features of using smoothed nonparametric estimation, we propose a novel test based on the Discrete Fourier Transform.

Throughout this chapter, i denotes the imaginary number such that $i = \sqrt{-1}$, A^* denotes the complex conjugate of A, $Re(A)$ means the real part of A, and $\|\cdot\|$ denotes the Euclidean norm. We further let $X_t(u) \equiv X_t e^{iu2\pi t/T}$ be the product of a random variable X_t and the Fourier basis function of time, where $X_t(0) = X_t$. Furthermore, we let $\hat{Q}_{xx}(u) \equiv \frac{1}{T}\sum_{t=1}^{T} X_t(u)X_t'$ be a complex-valued random matrix of u that depends on sample size T. When $u = 0$, we have $\hat{Q}_{xx}(0) \equiv \hat{Q}_{xx} = \frac{1}{T}\sum_{t=1}^{T} X_t X_t'$. And we let $Q_{xx}(u) \equiv \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E(X_t X_t')e^{iu2\pi t/T}$ be the probability limit for all $u \in \mathbb{R}$. In particular, when X_t is weakly stationary, $Q_{xx}(u) = E(X_t X_t')\int_0^1 e^{iu2\pi\tau}d\tau$. Here we allow X_t to be nonstationary such that it can change smoothly or shift abruptly. The nonstationarity here is similar to the asymptotical mse-stationarity in Hansen (2000).

Definition 3.1.1 (Hansen, 2000). *An array $\{X_t\}_{t=1}^{T}$ is asymptotically mse-stationary if*

$$\frac{1}{T}\sum_{t=1}^{[\tau T]} X_t X_t' \xrightarrow{p} \tau Q_{xx},$$

where $Q_{xx} = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E(X_t X_t')$.

The asymptotical mse-stationarity allows for certain deterministic cyclic process, seasonal dummies, and 'stationary' forms of heteroskedasticity. In this essay, our definition of nonstationarity is weaker than the asymptotical mse-stationarity since we do not need the weak convergence to hold for all $\tau \in [0, 1]$. Our analysis does not involve any sample splitting, and we only require the weak convergence to hold in the whole sample, i.e., $\tau = 1$. Therefore, we can allow X_t to have structural breaks of unknown forms, and that is an improvement over the literature.

The idea of our test is to capture the time-varying feature of β_t without estimating it directly. However, β_t is incorporated in Y_t through X_t. Given we do not restrict X_t to be stationary, the source of structural changes in Y_t could be structural changes in the marginal distribution of X_t. So we need first to purge the irrelevant information from Y_t via the OLS.

Consider the OLS estimator $\hat{\beta}$:

$$\begin{aligned}\hat{\beta} &= \hat{Q}_{xx}^{-1}\frac{1}{T}\sum_{t=1}^{T} X_t Y_t \\ &= \hat{Q}_{xx}^{-1}\frac{1}{T}\sum_{t=1}^{T} X_t X_t' \beta_t + \hat{Q}_{xx}^{-1}\frac{1}{T}\sum_{t=1}^{T} X_t \varepsilon_t \\ &\xrightarrow{p} \tilde{\beta}\end{aligned}$$

where

$$\tilde{\beta} \equiv Q_{xx}^{-1} \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E(X_t X_t')\beta_t$$

is a weighted average of β_t. In particular, when X_t is weakly stationary, $\tilde{\beta} = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \beta_t$. We can see the probability limit of the OLS estimator $\hat{\beta}$ is a constant vector with respect to time. As a result, it cannot capture the time-varying feature of β_t at each time t and such information will be contained in the estimated residuals. Consider the following decomposition of $\hat{\varepsilon}_t$:

$$\hat{\varepsilon}_t = Y_t - X_t'\hat{\beta}$$

$$
\begin{aligned}
&= X_t'(\beta_t - \hat{\beta}) + \varepsilon_t \\
&= \varepsilon_t - X_t'\hat{Q}_{xx}^{-1}\left(\frac{1}{T}\sum_{t=1}^{T} X_t\varepsilon_t\right) + X_t'\left[\beta_t - \hat{Q}_{xx}^{-1}\left(\frac{1}{T}\sum_{t=1}^{T} X_tX_t'\beta_t\right)\right].
\end{aligned}
$$

As can be seen, $\hat{\varepsilon}_t$ is decomposed into three components. The first element is the model disturbance ε_t. The second element is the estimation uncertainty that will vanish to 0 in probability as the sample size $T \to \infty$. The last is a weighted difference between β_t and a constant vector that does not depend on t. Under $\mathbb{H}_A$, estimated residual contain the time-varying feature of β_t. Using the Fourier transform, we can extract β_t's local feature over time without estimating it directly.

Consider the following complex-valued empirical process:

$$
\hat{A}(u) = \frac{1}{T}\sum_{t=1}^{T} X_t\hat{\varepsilon}_t e^{iu2\pi t/T}, \tag{3.2}
$$

which is the Discrete Fourier Transform (DFT) of $X_t\hat{\varepsilon}_t$. By the definition of $\hat{\varepsilon}_t$, we can further rewrite it as

$$
\begin{aligned}
\hat{A}(u) &= \frac{1}{T}\sum_{t=1}^{T} X_t\hat{\varepsilon}_t e^{iu2\pi t/T} \\
&= \frac{1}{T}\sum_{t=1}^{T}\left[X_t(u) - \hat{Q}_{xx}(u)\hat{Q}_{xx}^{-1}X_t\right](X_t'\beta_t + \varepsilon_t) \\
&= \frac{1}{T}\sum_{t=1}^{T}\hat{M}_t(u)X_t'\beta_t + \frac{1}{T}\sum_{t=1}^{T}\hat{M}_t(u)\varepsilon_t \\
&\equiv \hat{A}_1(u) + \hat{A}_2(u),
\end{aligned}
$$

where we let

$$
\hat{M}_t(u) = X_t e^{iu2\pi t/T} - \hat{Q}_{xx}(u)\hat{Q}_{xx}^{-1}X_t
$$

be a $d \times 1$ complex-valued random function of u and it has the following property:

$$
\frac{1}{T}\sum_{t=1}^{T}\hat{M}_t(u)X_t' = 0 \tag{3.3}
$$

for all $u \in \mathbb{R}$. $\hat{M}_t(u)$ can be viewed as a projection of X_t to a frequency space which is orthogonal to X_t under $\mathbb{H}_0$. Hence our test is robust to unknown structural changes in X_t. Intuitively, $X_t\hat{\varepsilon}_t$ contains the true model disturbance and estimation uncertainty and they will converge to 0 regardless of whether X_t is stationary of not.

When there is no structural change, $\hat{A}_1(u)$ is identically 0 for all u. Then the DFT is dominated by $\hat{A}_2(u)$ which is purely a stochastic process that converges to 0 in frequency domain (zero spectrum). While under $\mathbb{H}_A$, $\hat{M}_t(u)$ is no longer orthogonal to $X_t'\beta_t$. Then $\hat{A}_1(u)$ will capture the time-varying property of both X_t and β_t. As a result, the DFT is dominated by $\hat{A}_1(u)$, and it will converge to a nonzero spectrum in the frequency domain. In particular, if X_t is weakly stationary, and $\beta_t \equiv \beta(\frac{t}{T})$ is a smooth function of rescaled time $\frac{t}{T} \in (0, 1]$, then the probability limit of $\hat{A}_1(u)$ is proportional to the pseudo-covariance of β_t and $e^{iu2\pi t/T}$ in the sense that $\frac{t}{T}$ follows a standard uniform distribution on $[0, 1]$. Therefore, we can infer the existence of structural change by checking the different behaviors of the DFT at each frequency. Intuitively, the DFT is equivalent to running a nonparametric regression of Y_t on time because the Fourier basis function is an infinite sum of polynomials. Under $\mathbb{H}_0$, the DFT only captures the noise, and the impact of X_t has been purged. However, under $\mathbb{H}_A$, the DFT can capture the time-varying feature of both X_t and β_t and thus will converge to a nonzero spectrum.

The DFT also has an alternative theoretical interpretation such that it can be viewed as a generalized Hausman's test in frequency domain. We can show that the DFT is proportional to the difference between two estimators which converge to the same limit only under $\mathbb{H}_0$.

$$
\begin{aligned}
\hat{A}(u) &= \frac{1}{T}\sum_{t=1}^{T} X_t \hat{\varepsilon}_t e^{iu2\pi t/T} \\
&= \frac{1}{T}\sum_{t=1}^{T} X_t \left[Y_t - X_t'\hat{\beta}\right] e^{iu2\pi t/T} \\
&= \frac{1}{T}\sum_{t=1}^{T} X_t(u) Y_t - \hat{Q}_{xx}(u)\hat{Q}_{xx}^{-1}\frac{1}{T}\sum_{t=1}^{T} X_t Y_t \\
&= \hat{Q}_{xx}(u)\left[\hat{Q}_{xx}^{-}(u)\frac{1}{T}\sum_{t=1}^{T} X_t(u) Y_t - \hat{Q}_{xx}^{-1}\frac{1}{T}\sum_{t=1}^{T} X_t Y_t\right] \\
&\equiv \hat{Q}_{xx}(u)[\hat{\beta}(u) - \hat{\beta}],
\end{aligned}
$$

where $\hat{\beta}(u) \equiv \hat{Q}_{xx}(u)^{-}\frac{1}{T}\sum_{t=1}^{T} X_t(u)Y_t$ is a complexed-valued random vector function of u, and $\hat{Q}_{xx}^{-}(u)$ denotes the generalized inverse[1] of $\hat{Q}_{xx}(u)$ at each u. We can show that $\hat{\beta}(u)$ is the solution to

$$\hat{\beta}(u) = \arg\min_{\beta\in\mathbb{R}^d} WSSR(\beta, u),$$

where $WSSR(\beta, u) \equiv \sum_{t=1}^{T}(Y_t - X_t'\beta)^2 e^{iu2\pi t/T}$ is a weighted sum of squared residuals. In matrix notation, let $\mathcal{X} = [X_1', X_2', ..., X_T']'$, $\mathcal{Y} = [Y_1, Y_2, ..., Y_T]'$, and $\mathcal{W}(u)$ be a diagonal matrix with the diagonal elements being $[e^{iu2\pi/T}, e^{iu2\pi 2/T}, ..., e^{iu2\pi T/T}]$, then $\hat{\beta}(u) = [\mathcal{X}'\mathcal{W}(u)\mathcal{X}]^{-}\mathcal{X}'\mathcal{W}(u)\mathcal{Y}$ is indeed a WLS (Weighted Least Squares) estimator. The weighting matrix assigns weights to each time t and the weights are determined by the Fourier basis functions of time t. The objective function $WSSR(\beta, u)$ can be viewed as a DFT of the squared residuals for each $\beta \in \Theta$ and $u \in \mathbb{R}$. Intuitively, when the unknown parameter is constant with respect to time, the weights $e^{iu2\pi t/T}$ will have no impact on the minimization problem. Therefore, the optimal solution to minimizing $WSSR(\beta, u)$ at each u should be the same. We show under $\mathbb{H}_0 : \beta_t = \beta_0$, $\hat{\beta}(u)$ is consistent for β_0:

$$\hat{\beta}(u) = \hat{Q}_{xx}(u)^{-1}\frac{1}{T}\sum_{t=1}^{T} X_t(u) Y_t$$

[1] Here we use the generalized inverse because $\sum_{t=1}^{T} e^{iu2\pi t/T} \to 0$ when $u \in \mathbb{Z}$.

$$
\begin{aligned}
&= \hat{Q}_{xx}(u)^{-1}\frac{1}{T}\sum_{t=1}^{T} X_t(u)(X_t'\beta_0 + \varepsilon_t) \\
&= \beta_0 + \hat{Q}_{xx}(u)^{-1}\frac{1}{T}\sum_{t=1}^{T} X_t(u)\varepsilon_t \\
&\xrightarrow{p} \beta_0,
\end{aligned}
$$

for each $u \in \mathbb{R} \setminus \mathbb{Z}$ given $E(\varepsilon_t|X_t) = 0$. As can be seen, although $\hat{\beta}(u)$ is a complex-valued function of the nuisance parameter u, the entire complex-valued term vanishes to 0 as $T \to \infty$ under $\mathbb{H}_0$. Therefore the probability limit of the WLS estimator at each $u \in \mathbb{R} \setminus \mathbb{Z}$ will always be identical to each other. However, if there exists structural change in the unknown parameter, the weights will capture that information since different u delivers different weights to each time point t. Thus, under $\mathbb{H}_A : \beta_t \neq \beta_0$, for each fixed $u \in \mathbb{R} \setminus \mathbb{Z}$,

$$
\begin{aligned}
\hat{\beta}(u) &= \hat{Q}_{xx}^{-1}(u)\frac{1}{T}\sum_{t=1}^{T} X_t(u)Y_t \\
&= \hat{Q}_{xx}^{-1}(u)\frac{1}{T}\sum_{t=1}^{T} X_t(u)X_t'\beta_t + \hat{Q}_{xx}^{-1}(u)\frac{1}{T}\sum_{t=1}^{T} X_t(u)\varepsilon_t \\
&\xrightarrow{p} Q_{xx}^{-1}(u)\lim_{T\to\infty}\sum_{t=1}^{T} E(X_tX_t')e^{iu2\pi t/T}\beta_t,
\end{aligned}
$$

where the probability limit of the WLS estimator is a complex-valued functions of u. It is different from the probability limit of the OLS $\hat{\beta}$, which is a flat spectrum in the frequency domain. The DFT is equivalent to comparing the difference between two estimators that converge to the true parameter under $\mathbb{H}_0$ but different limits under $\mathbb{H}_A$. In this regard, our test is a generalized Hausman's test in frequency domain. To ensure the consistency of our test, we need to check the distance between the WLS estimator $\hat{\beta}(u)$ and the OLS estimator $\hat{\beta}$ at each $u \in \mathbb{R}$. This can be achieved by applying certain functionals of the DFT $\hat{A}(u)$. In contrast to the generalized Hausman's test as in Chen and Hong (2012) that compares the difference between two estimators at each time point t, our comparison is made at each frequency u.

In fact, there has been a long history of using spectral analysis in time series, e.g., see Granger and Hatanaka (1964), Hannan (1965, 1967), Engle (1974), Granger and Watson (1984), Choi and Phillips (1993), and Corbae et al. (2002). Among many others, Engle (1974) proposes a "Band Spectrum Regression" for static time series models. The idea is first to transform X_t and Y_t from the time domain to frequency domain using the DFT and then run an OLS using the corresponding DFTs. The sample is no longer indexed by time but frequency. Then we can get the same estimator as the conventional OLS in time domain. By running regressions at different frequencies, we can get insights on whether the model can capture various aspects of the data. Unlike the literature that mainly focuses on spectrum analysis in stationary time series, this chapter applies frequency domain analysis to on nonstationary time series, namely analysis of structural change.

The DFT projects the time-varying feature of β_t to the frequency domain without loss of information. By this device, we can avoid the smoothed nonparametric estimation of β_t and can test $\mathbb{H}_0$ consistently by checking each frequency. Our test does not need prior information about the type of structural change. As long as β_t is a non-constant function of time t, the DFT will capture that information and $\hat{A}_1(u)$ will converge to a nonzero spectrum because $\hat{\beta}(u)$ and $\hat{\beta}(0)$ converge to different probability limits. Most importantly, our test can detect a class of local alternatives that converges to $\mathbb{H}_0$ at a parametric rate $T^{-1/2}$. Therefore, our test is asymptotically more powerful than those of Chen and Hong (2012), Zhang and Wu (2012), and Cai et al. (2015), in which the rate of local alternatives that can be detected is at $T^{-1/2}h^{-1/4}$.

3.2 Asymptotic Theory

To determine whether the DFT is significantly different from a zero spectrum, we need to provide the asymptotic theory for the test based on the Fourier transform. Let "$\xrightarrow{p}$", "$\xrightarrow{d}$", and "$\Rightarrow$"denote convergence in probability, convergence in distribution, and weak convergence respectively. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. For any two sigma-fields $\mathcal{A}$ and $\mathcal{B}$, define the following absolute regular mixing coefficient

$$\alpha(\mathcal{A}, \mathcal{B}) = \frac{1}{2} \sup \sum_{j=1}^{J} \sum_{k=1}^{K} \left| P(A_j \cap B_k) - P(A_j)P(B_k) \right|,$$

where $A_i \in \mathcal{A}$, $B_j \in \mathcal{B}$, and the supremum is taken over all pairs of finite partitions $\{A_j\}_{j \in J}$ and $\{B_k\}_{k \in K}$ of the sample space that are $\mathcal{A}$ and $\mathcal{B}$ measurable respectively. The absolute regularity condition is necessary for us to establish the uniform convergence between the sample moments to the population moments of the complexed-valued process. However, since we do not restrict $\{X_t', \varepsilon_t\}'$ to be stationary, we need a stronger condition to restrict the temporal dependence to be uniformly weak. Let $-\infty \leq J_1 \leq J_2 \leq \infty$, and $\mathcal{F}_{J_1}^{J_2} \equiv \sigma(X_t, \varepsilon_t)$ such that $J_1 \leq t \leq J_2$, then for each $m \geq 1$, define the following dependence coefficient:

$$\alpha(m) = \sup_{t \in \mathbb{Z}} \alpha(\mathcal{F}_{-\infty}^{t}, \mathcal{F}_{t+m}^{\infty}).$$

We define mixing coefficient $\alpha(m)$ as in Bradley (2005) for nonstationary random sequences. Because the distribution at each t could be different, we have to take supremum over all $t \in \mathbb{Z}$ to restrict the largest possible dependence. When $\{X_t', \varepsilon_t\}'$ is strictly stationary, one simply has

$$\alpha(m) = \alpha(\mathcal{F}_{-\infty}^{0}, \mathcal{F}_{m}^{\infty}).$$

Now we provide the following regularity conditions.

Assumption 3.2.1. *$\{X_t', \varepsilon_t\}_{t=1}^{\prime T}$ is a $(d+1)\times 1$ absolutely regular process uniformly over $t \in \mathbb{Z}$ with the mixing coefficient such that $\sum_{j=1}^{\infty} \alpha(j)^{\frac{\delta-1}{\delta}} < C < \infty$ for some $\delta > 1$.*

Assumption 3.2.2. *$\{\varepsilon_t\}$ is a MDS process such that $E(\varepsilon_t | I_{t-1}) = 0$ a.s. for all t, where I_{t-1} is a sigma-field generated by $\{X_{t-1}X_{t-2}, ..., \varepsilon_{t-1}, \varepsilon_{t-2}, ...\}$.*

Assumption 3.2.3. *$E(\varepsilon_t | X_t) = 0$ almost surely for all t.*

Assumption 3.2.4. *(i) $V_{xx}(u_1, u_2) \equiv \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} Re\left\{E[X_t(u_1)X_t(u_2)^* \varepsilon_t^2]\right\}$ is finite and positive definite for all $(u_1, u_2) \in \mathbb{R}^2$; (ii) $Q_{xx} = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} E(X_t X_t')$ is non-singular and finite.*

Assumption 3.2.5. *$E(X_{jt}^{4\delta}) \leq C < \infty$ for all $j = 1, ..., d$ and all t; $E(\varepsilon_t^{4\delta}) \leq C < \infty$ for all t.*

Assumption 3.2.6. *$\beta_t \in \Theta$, where Θ is a bounded parameter space in $\mathbb{R}^d$ such that $\frac{1}{T} \sum_{t=1}^{T} ||\beta_t||^2 < C < \infty$.*

Assumption 3.2.1 allows the structure of X_t to change over time but restricts its temporal dependence to be weak. With this condition, our test applies to dynamic time series regression models. The structural change in X_t can be either smooth or abrupt. That means our test is robust to unknown structural change in the covariates. The mixing condition restricts the temporal dependence to be weak uniformly over time t. This is crucial for us to apply the law of large numbers and the central limit theorem by Bradley and Tone (2015). We note that Assumption 3.2.1 excludes the case when X_t is unit root or near unit root. Such cases will be investigated in subsequent studies.

Assumption 3.2.2 restricts the model disturbance ε_t to be serially uncorrelated. This is only for the simplicity of delivering our main result. We can allow ε_t to exhibit serial correlation and use the long-run variance-covariance (e.g., Newey and West, 1987). Assumption 3.2.3 excludes the existence of endogenous covariates. In Section 3.3, we will consider endogenous covariates. Assumption 3.2.3 implies the mean of ε_t is 0 and does not vary with time, although the structural of X_t may vary with time. Assumption 3.2.4(i) allows ε_t to exhibit conditional heteroskedasticity which is common for time series data. Assumption 3.2.4(ii) ensures the existence of the OLS. Assumption 3.2.5 is a regular moment condition on ε_t and X_t. Assumption 3.2.5 implies the variance of ε_t can be time-varying. If both X_t and ε_t are weakly stationary, then Assumption 3.2.4(i) can be implied by Assumption 3.2.5. Compared to the existing literature, e.g., Chen and Hong (2012), Kristensen (2012), Cai et al. (2015), we do not impose any smoothness assumption on the unknown parameter β_t. We only need it to be L-2 bounded. This is imposed in Assumption 3.2.6. Our assumptions are quite weak.

Let

$$\hat{V}_{xx}(u_1, u_2) \equiv \frac{1}{T} \sum_{t=1}^{T} Re[X_t(u_1) X_t(u_2)^*] \varepsilon_t^2,$$

be the sample analog for $V_{xx}(u_1, u_2)$.

Lemma 3.2.1. *Suppose Assumptions 3.2.1 to 3.2.5 hold, then as $T \to \infty$*

$$\sup_{u \in \mathbb{R}} \left\| \hat{Q}_{xx}(u) - Q_{xx}(u) \right\| \xrightarrow{p} 0,$$

$$\sup_{u_1, u_2 \in \mathbb{R}^2} \left\| \hat{V}_{xx}(u_1, u_2) - V_{xx}(u_1, u_2) \right\| \xrightarrow{p} 0,$$

for all $u \in \mathbb{R}$.

Lemma 3.2.1 provides the uniform convergence that is necessary for our asymptotic result. The proof is quite straightforward given the complex-valued function $e^{iu2\pi t/T}$ is always bounded by 1 for all $u \in \mathbb{R}$. Thus, the uniform convergence can be shown accordingly.

Theorem 3.2.1. *Suppose Assumptions 3.2.1-3.2.5 hold, under $\mathbb{H}_0$,*

$$\sqrt{T}\hat{A}(u) \Rightarrow \mathcal{G}(u),$$

where $\mathcal{G}(u)$ is a zero-mean complex-valued Gaussian process with covariance kernel

$$\begin{aligned}
\mathcal{K}(u_1, u_2) &\equiv cov[\mathcal{G}(u_1), \mathcal{G}(u_2)^*] \\
&= V_{xx}(u_1, u_2) - Q_{xx}(u_1)Q_{xx}^{-1}(0)V_{xx}(0, u_2) - V_{xx}(u_1, 0)Q_{xx}^{-1}(0)Q_{xx}(u_2)^* \\
&\quad +Q_{xx}(u_1)Q_{xx}^{-1}(0)V_{xx}(0, 0)Q_{xx}^{-1}(0)Q_{xx}(u_2)^*.
\end{aligned}$$

Theorem 3.2.1 provides the asymptotic null distribution of the DFT $\hat{A}(u)$ scaled by $\sqrt{T}$. At each frequency u, it converges to a normal distribution with mean 0 and variance $\mathcal{K}(u, u)$ as $T \to \infty$. If both X_t and ε_t are weakly stationary, then

$$Q_{xx}(u) = E(X_t X_t') \int_0^1 e^{iu2\pi\tau} d\tau,$$

and

$$V_{xx}(u_1, u_2) = E(X_t X_t \varepsilon_t^2) \int_0^1 e^{i2\pi(u_1 - u_2)\tau} d\tau.$$

We can show that

$$\mathcal{K}(u_1, u_2) = E(X_t X_t \varepsilon_t^2)\widetilde{cov}(e^{iu_1 2\pi\tau}, e^{-iu_2 2\pi\tau}),$$

where $\widetilde{cov}(e^{iu_1 2\pi\tau}, e^{-iu_2 2\pi\tau}) = \left[\int_0^1 e^{i2\pi(u_1-u_2)\tau} d\tau - \int_0^1 \int_0^1 e^{i2\pi(u_1\tau - u_2\eta)} d\tau d\eta\right]$ is a pseudo-covariance in the sense that τ follows the $U[0, 1]$ distribution. Intuitively, the covariance kernel contains two components. $E(X_t X_t \varepsilon_t^2)$ is the asymptotic variance of $\frac{1}{\sqrt{T}} \sum_{t=1}^T X_t \varepsilon_t$, which is the uncertainty introduced by the OLS estimation. The other component can be viewed as the noise introduced by the Fourier transform. When X_t and ε_t are nonstationary, the two components will intertwine with each other.

Next, we want to show that the DFT $\hat{A}(u)$ will converge to a non-zero spectrum under $\mathbb{H}_A$ in the frequency domain. This is essential to ensure the consistency of our test.

Theorem 3.2.2. *Suppose Assumptions 3.2.1 to 3.2.6 hold, under $\mathbb{H}_A$, as $T \to \infty$*

$$\sup_{u \in \mathbb{R}} \left\| \hat{A}(u) - \tilde{A}(u) \right\| \xrightarrow{p} 0,$$

where $\tilde{A}(u) \equiv \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \left[I_d e^{iu2\pi t/T} - Q_{xx}(u) Q_{xx}^{-1} \right] E(X_t X_t') \beta_t$.

Theorem 3.2.2 gives the probability limit of $\hat{A}(u)$ which is a non-flat spectrum in the frequency domain. If we let X_t be weakly stationary and $\beta_t \equiv \beta(\frac{t}{T})$ be a smooth function of rescaled time $t/T \in (0, 1]$, then

$$\begin{aligned} \tilde{A}(u) &= E(X_t X_t') \lim_{T\to\infty} \left[\frac{1}{T} \sum_{t=1}^{T} \beta_t e^{iu2\pi t/T} - \frac{1}{T} \sum_{t=1}^{T} \beta_t \frac{1}{T} \sum_{t=1}^{T} e^{iu2\pi t/T} \right] \\ &\equiv E(X_t X_t') \widetilde{cov}[\beta(\tau), e^{iu2\pi\tau}], \end{aligned}$$

where $\widetilde{cov}[\beta(\tau), e^{iu2\pi\tau}] = \int_0^1 \beta(\tau) e^{iu2\pi\tau} d\tau - \int_0^1 \beta(\tau) d\tau \int_0^1 e^{iu2\pi\tau} d\tau$ is a pseudo-covariance in the sense that τ follows the $U[0, 1]$ distribution. As long as $\beta_t \equiv \beta(\frac{t}{T})$ is a non-constant function of time t, $\widetilde{cov}[\beta(\tau), e^{iu2\pi\tau}]$ will be different from 0 for $u \neq 0$. Theorem 3.2.2 shows that the DFT $\hat{A}(u)$ is equivalent to the sample analog of a pseudo-covariance between β_t and $e^{iu2\pi t/T}$. The result also explains why we construct $\hat{A}(u)$ by considering the DFT of $X_t \hat{\varepsilon}_t$ rather than $\hat{\varepsilon}_t$. When X_t is weakly stationary such that $E(X_t) = 0$, then our test will have no power if we consider the Fourier transform of $\hat{\varepsilon}_t$. For this reason, we construct $\hat{A}(u)$ using the Fourier transform of $X_t \hat{\varepsilon}_t$ to ensure that it converges to a nonzero spectrum under $\mathbb{H}_A$.

One way to examine the behavior the DFT in the frequency domain is to compose a test for each fixed u. However, the power of a test using $\hat{A}(u)$ depends on u under $\mathbb{H}_A$. Intuitively, $\hat{A}(u)$ can be viewed as a series approximation of a weighted average of β_t using the Fourier basis functions. Frequency u represents how volatile the $sin(\cdot)$ and $cos(\cdot)$ series are. If the variation of β_t over time is very slow, then we expect the Fourier basis functions with a lower frequency can approximate β_t better than those with a higher frequency. Thus, we should test $\mathbb{H}_0$ by assigning large weights to small frequencies, and *vice versa*. Unfortunately, the information on β_t is usually unknown *a prior*. To ensure consistency against all alternatives, we consider the following Kolmogorov-Smirnov (KS) and Cramér-von Mises (CM) type test statistics:

$$\hat{K} = T \sup_{u \in \mathbb{R}} ||\hat{A}(u)||^2,$$

and

$$\hat{C} = T \int_{\mathbb{R}} ||\hat{A}(u)||^2 W(u) du,$$

where $W(\cdot)$ is a weighting function that assigns weights on each frequency, and it satisfies

$$\int_{R} |u|^2 W(u) du < \infty. \tag{3.4}$$

$||\hat{A}(u)||^2$ is called the 'periodogram and it reveals the magnitude of the DFT at each frequency. The periodogram is also proportional to the quadratic difference between the WLS and the OLS estimators at each frequency.

The computation of $\hat{C}$ involves integration over all u. We note that proper choices of the weighting function $W(u)$ can avoid numerical integration of u, and $\hat{C}$ will have a closed-form expression. One example is the zero-mean normal density function

$$W(u) = \frac{1}{\sqrt{2\pi\xi^2}} exp\left(\frac{u^2}{2\xi^2}\right).$$

By the following identity

$$\int_{\mathbb{R}} cos(uz) \frac{1}{\sqrt{2\pi\xi^2}} exp\left(-\frac{u^2}{2\xi^2}\right) du = exp\left(-\frac{z^2\xi^2}{2}\right),$$

we have

$$\begin{aligned}\hat{C} &= T\int_{\mathbb{R}} \hat{A}(u)'\hat{A}(u)^{*}W(u)du \\ &= \frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T} X_s'X_t\hat{\varepsilon}_s\hat{\varepsilon}_t exp\left(-\frac{2\pi^2(s-t)^2\xi^2}{T^2}\right),\end{aligned}$$

where ξ is the standard deviation that measures the dispersion of weights assigned around 0. In general, the results are insensitive to the choice of ξ. Another example is the Laplace density weighting

$$W(u) = \frac{\lambda}{2}e^{-\lambda|u|}.$$

By the identity that

$$\int_{\mathbb{R}} cos(uz)\frac{\lambda}{2}e^{-\lambda|u|} = \frac{\lambda^2}{\lambda^2+z^2},$$

we can show

$$\begin{aligned}\hat{C} &= T\int_{\mathbb{R}} \hat{A}(u)'\hat{A}(u)^{*}W(u)du \\ &= \frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T} X_s'X_t\hat{\varepsilon}_s\hat{\varepsilon}_t \left(\frac{\lambda^2T^2}{\lambda^2T^2+4\pi^2(s-t)^2}\right).\end{aligned}$$

Here λ is the parameter that measures the dispersion of weights assigned around 0. As λ increases, more weights are assigned to higher frequencies.

After providing the test statistics $\hat{K}$ and $\hat{C}$ based on the DFT, we now show the asymptotic null distributions.

Theorem 3.2.3. *Suppose Assumptions 3.2.1-3.2.5 hold, under $\mathbb{H}_0$, by the continuous mapping theorem*

$$\hat{K} \xrightarrow{d} \sup_{u\in\mathbb{R}} ||\mathcal{G}(u)||^2,$$

and

$$\hat{C} \xrightarrow{d} \int_{\mathbb{R}} ||\mathcal{G}(u)||^2 W(u)du,$$

as $T \to \infty$.

Theorem 3.2.3 provides the asymptotic distributions of $\hat{K}$ and $\hat{C}$ under $\mathbb{H}_0$. Our test is tuning parameter free, and the asymptotic distributions are not pivotal because they depend on the unknown data generating process (DGP). Thus, resampling methods are needed to obtain critical values. We use the following resampling method proposed by Hansen (1996).

- Step(i). Use the sample $\{Y_t, X_t'\}_{t=1}^T$ to estimate the model via the OLS, and compute $\hat{K}$ and $\hat{C}$;
- Step (ii). Generate $i.i.d.N(0,1)$ random variables $\{v_{bt}\}_{t=1}^T$ and compute compute $\hat{K}_b$ and $\hat{C}_b$ using $\hat{A}_b(u) = \frac{1}{\sqrt{T}} \sum_{t=1}^T \hat{M}_t(u)\hat{\varepsilon}_t v_{bt}$, where

  $$\hat{M}_t(u) = X_t e^{iu2\pi t/T} - \hat{Q}_{xx}(u)\hat{Q}_{xx}^{-1} X_t,$$

  and $\hat{\varepsilon}_t$ is the estimated residuals from OLS;
- Step (iii). Repeat Step (ii) for a total of B times to obtain B bootstrap test statistics $\{\hat{K}_b, \hat{C}_b\}_{b=1}^B$;
- Step (iv). Compute the bootstrap p-values for $\hat{K}$ and $\hat{C}$ respectively, with $p_{B,T}^K = \frac{1}{B}\sum_{b=1}^B \mathbf{1}(\hat{K}_b > \hat{K})$ and $p_{B,T}^C = \frac{1}{B}\sum_{b=1}^B \mathbf{1}(\hat{C}_b > \hat{C})$, where $\mathbf{1}(\cdot)$ is the indicator function.

$p_{B,T}^K$ and $p_{B,T}^C$ are the conditional p-values for $\hat{K}$ and $\hat{C}$ via resampling for B times. Next, we show they are consistent for the true p-values. Let $F^K(\cdot)$ and $F^C(\cdot)$ denote the distribution of function of $\sup_{u\in\mathbb{R}} ||\mathcal{G}(u)||^2$ and $\int_{\mathbb{R}} ||\mathcal{G}(u)||^2 W(u)du$ respectively. Define $p_T^K = 1 - F^K(\hat{K})$ and $p_T^C = 1 - F^C(\hat{C})$ be the asymptotic p-values, such that

$$\lim_{T\to\infty} Pr\{p_T^j \leq \alpha | \mathbb{H}_0\} = \alpha,$$

for $j = K, C$. Conditional on the sample $\{Y_t, X_t\}_{t=1}^T$, let $F_T^K(\cdot)$ and $F_T^C(\cdot)$ be the conditional distribution functions of $\hat{K}$ and $\hat{C}$, respectively. Define $\tilde{p}_T^K = 1 - F_T^K(\hat{K})$

and $\tilde{p}_T^C = 1 - F_T^C(\hat{C})$ to be the conditional p-values, where $\hat{F}_T^K(\cdot)$ and $\hat{F}_T^K(\cdot)$ are generated by replacing $\hat{A}(u)$ with

$$\hat{A}(u) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \hat{M}_t(u)\hat{\varepsilon}_t v_t,$$

where $\{v_t\}_{t=1}^T$ is $i.i.d.N(0,1)$. Then we can show the consistency of the resampling method by the following Corollary.

Corollary 3.2.1. *Under the regularity conditions in Theorem 3.1, $p_{B,T}^j \xrightarrow{p} \tilde{p}_T^j$ as $B \to \infty$, and*

$$\tilde{p}_T^j = p_T^j + op(1),$$

for $j = K, C$. Therefore

$$\tilde{p}_T^j \Rightarrow p^j,$$

where p^j is the true p-value for $j = K, C$.

Unlike Andrews (1993), Andrews and Ploberger (1994), and Bai and Perron (1998), in which nuisance parameter has to be a subset of $(0,1)$, we do not have to restrict u to be a subset of $\mathbb{R}$. By using frequency domain based method, our test can detect structural breaks near the boundary regions.

Next, we show the asymptotic power of our test statistics.

Theorem 3.2.4. *Under Assumptions 3.2.1-3.2.6 and $\mathbb{H}_A$, for any sequence of non-stochastic constants $\{c_T = o(T)\}$, as $T \to \infty$,*

$$\begin{aligned} P(\hat{K} > c_T) &\to 1, \\ P(\hat{C} > c_T) &\to 1. \end{aligned}$$

This result shows the power of $\hat{K}$ and $\hat{C}$ against fixed alternatives approaches to 1 as $T \to \infty$. Therefore, the tests $\hat{K}$ and $\hat{C}$ are consistent against both abrupt structural breaks and smooth structural changes. Compared to Chen and Hong (2012), Zhang and Wu (2012), and Cai et al. (2015), $\hat{K}$ and $\hat{C}$ avoid smoothed nonparametric estimation.

We now investigate the local power property of our tests and compare with existing consistent tests in the literature. Consider the following local alternatives under $\mathbb{H}_{A1}$:

$$\beta_t = \beta_0 + \delta_T \phi_t,$$

where $\phi_t \in \mathbb{R}^d$ is a nonrandom function of time t such that $0 < \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} ||\phi_t||^2 < \infty$. ϕ_t can be either a smooth function of $\frac{t}{T}$ (smooth structural changes) or a step function of t (abrupt structural breaks).

Theorem 3.2.5. *Suppose Assumptions 3.2.1-3.2.6 and $\mathbb{H}_{A1}$ with $\delta_T = T^{-1/2}$ hold. Then as $T \to \infty$,*

$$\begin{aligned}
\hat{K} &\xrightarrow{d} \sup_{u\in\mathbb{R}} ||\xi(u) + G(u)||^2, \\
\hat{C} &\xrightarrow{d} \int_{\mathbb{R}} ||\xi(u) + G(u)||^2 W(u) du,
\end{aligned}$$

where

$$\xi(u) = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \left[I_d e^{iu2\pi t/T} - Q_{xx}(u) Q_{xx}^{-1} \right] E(X_t X_t') \phi_t.$$

Theorem 3.2.5 shows that our tests $\hat{K}$ and $\hat{C}$ have nontrivial power against a class of local alternatives with the parametric rate $\delta_T = T^{-1/2}$. In contrast, Chen and Hong (2012), Zhang and Wu (2012), and Cai et al. (2015) can only detect a class of local alternatives at a rate of $\delta_T = T^{-1/2} h^{-1/4}$, where the bandwidth $h \to 0$ as $T \to \infty$. This is the advantage of using the Fourier transform by avoiding smoothed nonparametric estimation. The price we have to pay is that the asymptotic distributions of our tests $\hat{K}$ and $\hat{C}$ are not pivotal. Moreover, we are free from boundary problem which is common in smoothed nonparametric testing. Chen and Hong (2012) use boundary reflection to increase the convergence rate of boundary estimation. Our test does not require such a delicate treatment. Compared to popular tests that achieve parametric rates such as Andrews' (1993) supremum and Bai and Perron's (1998) double maximum tests, our test does not require trimming of the data. Their tests do not have uniform power for all $t/T \in (0, 1]$.

3.3 Extension to Endogenous Covariates

In this section, we extend our tests to a time series regression model with endogenous covariates, i.e., $E(\varepsilon_t|X_t) \neq 0$. Suppose there exists a set of instruments $Z_t \in \mathbb{R}^l$ such that $E(\varepsilon_t|Z_t) = 0$ a.s., and $E(X_t Z_t') \neq 0$, where $l \geq d$.

To test $\mathbb{H}_0$, we now consider the following complexed-valued empirical process:

$$\hat{A}^{IV}(u) \quad = \quad \frac{1}{T}\sum_{t=1}^{T} \hat{X}_t \hat{\varepsilon}_t e^{iu2\pi t/T},$$

where $\hat{\varepsilon}_t = Y_t - X_t'\hat{\beta}_{2sls}$ is the estimated residual from the Two-Stage Least Squares (2SLS) estimation, and $\hat{X}_t \in \mathbb{R}^d$ is the fitted value from the first stage regression of X_t on the $l \times 1$ dimensional instrumental variables Z_t. $\hat{A}^{IV}(u)$ can be viewed as the DFT of $\hat{X}_t\hat{\varepsilon}_t$. The 2SLS estimator cannot capture the time-varying feature of the unknown parameter. So we will focus on the estimated residual $\hat{\varepsilon}_t$ from the 2SLS and consider the following decomposition:

$$\hat{\varepsilon}_t \quad = \quad Y_t - X_t'\hat{\beta}_{2sls}$$

$$
\begin{aligned}
&= X_t'(\beta_t - \hat{\beta}_{2sls}) + \varepsilon_t \\
&= X_t'\left[\beta_t - \left(\hat{Q}_{xz}\hat{Q}_{zz}^{-1}\hat{Q}_{zx}\right)^{-1}\hat{Q}_{xz}\hat{Q}_{zz}^{-1}\hat{Q}_{zy}\right] + \varepsilon_t \\
&= \varepsilon_t - X_t'\left(\hat{Q}_{xz}\hat{Q}_{zz}^{-1}\hat{Q}_{zx}\right)^{-1}\hat{Q}_{xz}\hat{Q}_{zz}^{-1}\left(\frac{1}{T}\sum_{t=1}^{T} Z_t\varepsilon_t\right) \\
&\quad + X_t'\left[\beta_t - \left(\hat{Q}_{xz}\hat{Q}_{zz}^{-1}\hat{Q}_{zx}\right)^{-1}\hat{Q}_{xz}\hat{Q}_{zz}^{-1}\left(\frac{1}{T}\sum_{t=1}^{T} Z_t X_t'\beta_t\right)\right].
\end{aligned}
$$

Like in the case where all covariates are exogenous, the estimated residual from the 2SLS can also be decomposed into three components: the model disturbance ε_t, the estimation uncertainty, and the weighted difference between β_t and a constant vector that does not depend on t. Under $\mathbb{H}_0 : \beta_t = \beta_0$ for all t, $\hat{\beta}_{2sls}$ is consistent for β_0. Then the third component will be identically 0 and the DFT will converge to 0 in probability for all frequencies under the orthogonality condition between ε_t and Z_t. While under $\mathbb{H}_A$, $\hat{\beta}_{2sls}$ fails to capture the time-varying feature of β_t, and such information will be contained in the estimated residuals. Therefore, the DFT will converge to a nonzero spectrum in the frequency domain. In particular, if both X_t and Z_t are weakly stationary, and $\beta_t \equiv \beta(\frac{t}{T})$ is a smooth function of the rescaled time $\frac{t}{T} \in (0, 1]$, then the probability limit of the DFT under $\mathbb{H}_A$ is proportional to the pseudo-covariance of β_t and $e^{iu2\pi t/T}$ in the sense that $\frac{t}{T}$ follows a standard uniform distribution on $[0, 1]$.

For the first stage reduced form, we simply estimate it by the OLS:

$$\hat{X}_t = \hat{\gamma}' Z_t,$$

where $\hat{\gamma} = (\sum_{t=1}^{T} Z_t Z_t')^{-1} \sum_{t=1}^{T} Z_t X_t' \equiv \hat{Q}_{zz}^{-1}\hat{Q}_{zx}$ is the OLS estimator. As $T \to \infty$, $\hat{X}_t$ will converge in probability to $\tilde{X}_t \equiv \gamma' Z_t$. It is possible that the true relationship between X_t and Z_t is unstable such that γ is not constant with respect to time. If that is the case, then $\hat{\gamma}$ will not be consistent for γ. However, that has a trivial impact on our analysis since our goal is not to estimate γ consistently but rather to use Z_t as a proxy to obtain the 2SLS estimator. So $\tilde{X}_t$ can be viewed as a pure linear projection of X_t on Z_t. As long as X_t and Z_t are not orthogonal to each other, we can guarantee the consistency of the 2SLS estimators for β_0 under $\mathbb{H}_0$. Moreover, we do not restrict either X_t or Z_t to be weakly stationary. Thus $\tilde{X}_t$ can also be nonstationary such that its distribution may change smoothly or break abruptly. The fitted value $\hat{X}_t$ from the first stage regression serves as the regressor in the second stage. It has the same role as X_t in the case where all covariates are exogenous. Therefore its instability will have no impact on our testing for structural change in the structural equation. The tests by Hall et al. (2012) and Perron and Yamamoto (2014) rely on comparing consistent estimates for β_t using the 2SLS for subsamples. Therefore they need first to investigate the stability of the reduce form. This is a drawback of testing in the time domain. However, our frequency domain based approach can avoid this issue and is robust to instability in both covariates and instruments.

Like the DFT in Section 3.1, $\hat{A}^{IV}(u)$ can also be interpreted as a generalized Hausman's test in frequency domain. Let $\hat{X}_t(u) \equiv \hat{X}_t e^{iu2\pi t/T} = \hat{Q}_{xz}\hat{Q}_{zz}^{-1}Z_t e^{iu2\pi t/T}$, we can rewrite the DFT as

$$\begin{aligned}
\hat{A}^{IV}(u) &= \frac{1}{T}\sum_{t=1}^{T}\hat{X}_t(u)\hat{\varepsilon}_t \\
&= \hat{Q}_{\hat{x}x}(u)\left[\hat{\beta}^{IV}(u) - \hat{\beta}_{2sls}\right] \\
&= \hat{Q}_{xz}\hat{Q}_{zz}^{-1}\hat{Q}_{zx}(u)\left[\hat{\beta}^{IV}(u) - \hat{\beta}_{2sls}\right],
\end{aligned}$$

where we define the following linear IV estimator for β_0:

$$\begin{aligned}
\beta^{IV}(u) &\equiv \left[\hat{Q}_{xz}\hat{Q}_{zz}^{-1}\hat{Q}_{zx}(u)\right]^{-}\hat{Q}_{xz}\hat{Q}_{zz}^{-1}\frac{1}{T}\sum_{t=1}^{T}Z_tY_te^{iu2\pi t/T} \\
&= [\mathcal{X}'\mathcal{P}_Z\mathcal{W}(u)\mathcal{X}]^{-}\mathcal{X}'\mathcal{P}_Z\mathcal{W}(u)\mathcal{Y},
\end{aligned}$$

where $\mathcal{P}_Z = \mathcal{Z}(\mathcal{Z}'\mathcal{Z})^{-1}\mathcal{Z}'$ is a projection matrix. Compared to the WLS defined in Section 3.1, $\hat{\beta}^{IV}$ contains an additional projection to the space spanned by Z_t.

In particular, when $u = 0, \beta^{IV}(0) = [\mathcal{X}'\mathcal{P}_Z\mathcal{X}]^{-1}\mathcal{X}'\mathcal{P}_Z\mathcal{Y}$ is exactly the 2SLS estimator. We see the DFT $\hat{A}^{IV}(u)$ is proportional to the difference between two estimators $\hat{\beta}^{IV}(u)$ and $\hat{\beta}_{2sls}$. Under $\mathbb{H}_0$,

$$\begin{aligned}\beta^{IV}(u) &= \left[\hat{Q}_{xz}\hat{Q}_{zz}^{-1}\hat{Q}_{zx}(u)\right]^{-1}\hat{Q}_{xz}\hat{Q}_{zz}^{-1}\hat{Q}_{zy}(u) \\ &= \beta_0 + \left[\hat{Q}_{xz}\hat{Q}_{zz}^{-1}\hat{Q}_{zx}(u)\right]^{-1}\hat{Q}_{xz}\hat{Q}_{zz}^{-1}\frac{1}{T}\sum_{t=1}^{T} Z_t\varepsilon_t e^{iu2\pi t/T} \\ &\xrightarrow{p} \beta_0,\end{aligned}$$

for $u \in \mathbb{R} \setminus \mathbb{Z}$. Thus $\beta^{IV}(u)$ is consistent for β_0 under $\mathbb{H}_0$. $\hat{\beta}^{IV}(u)$ can be viewed as a general class of estimators that contain the 2SLS estimator as a special case. Following analogous arguments, $\hat{\beta}^{IV}(u)$ will converge to the same probability limit for all $u \in \mathbb{R}$ under $\mathbb{H}_0 : \beta_t = \beta_0$. Because the different weights to each observation have trivial impacts in large samples. However, $\hat{\beta}^{IV}(u)$ will converge to a nonzero spectrum under $\mathbb{H}_A$.

Now we provide the asymptotic theory for the test based on $\hat{A}^{IV}(u)$.

Assumption 3.3.1. *$\{X_t', Z_t', \varepsilon_t\}_{t=1}^{'T}$ is a $(d+l+1)\times 1$ absolutely regular process uniformly over $t \in \mathbb{Z}$ with the mixing coefficient such that $\sum_{j=1}^{\infty} \alpha(j)^{\frac{\delta-1}{\delta}} < C < \infty$ for some $\delta > 1$.*

Assumption 3.3.2. *$\{\varepsilon_t\}$ is a MDS process such that $E(\varepsilon_t|I_{t-1}) = 0$ a.s., where I_{t-1} is a Sigma-algebra generated by $\{Z_{t-1}, Z_{t-2}, ..., \varepsilon_{t-1}, \varepsilon_{t-2}, ...\}$*

Assumption 3.3.3. *$E(\varepsilon_t|X_t) \neq 0$ for some t and $E(Z_t\varepsilon_t) = 0$ a.s. for all t.*

Assumption 3.3.4. *(i) $V_{zz}(u_1, u_2) \equiv \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E[Z_t(u_1)Z_t(u_2)^*\varepsilon_t^2]$ is finite and positive definite for all $(u_1, u_2) \in \mathbb{R}^2$;*
(ii) $Q_{zz} = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E(Z_tZ_t')$ is nonsingular and finite;
(iii) $Q_{xz} = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E[X_tZ_t']$ is finite and of full rank.

Assumption 3.3.5. *$E(Z_{jt}^{4\delta}) \leq C < \infty$ and $E(X_{kt}^{4\delta}) \leq C < \infty$ for some $\delta > 1$, all $j = 1, ..., l$, $k = 1, ..., d$, and all t; $E(\varepsilon_t^{4\delta}) \leq C < \infty$ for all t.*

Like Assumption 3.2.1, Assumption 3.3.1 allows both X_t and Z_t to have smooth structural changes or abrupt structural breaks. The temporal dependence has been controlled by uniform mixing coefficient for possibly nonstationary time series. Such a restriction on the temporal dependence is widely used in time series analysis. Assumption 3.3.2 restricts the disturbance to be serially uncorrelated for the simplicity of deriving our result.[2] Assumption 3.3.3 characterizes the existence of endogeneity and the validity of instrumental variable Z_t. Assumptions 3.3.4 and 3.3.5 are regular moment conditions on X_t, Z_t, and ε_t which are rather mild.

By analogous proof as in Lemma 3.2.1, we can show the following

$$\sup_{u\in\mathbb{R}} \left\|\hat{Q}_{zx}(u) - Q_{zx}(u)\right\| \xrightarrow{p} 0,$$
$$\sup_{u_1,u_2\in\mathbb{R}^2} \left\|\hat{V}_{zz}(u_1,u_2) - V_{zz}(u_1,u_2)\right\| \xrightarrow{p} 0,$$

as $T \to \infty$, where $\hat{Q}_{zx}(u) = \frac{1}{T}\sum_{t=1}^{T} Z_t(u)X_t'$ and $\hat{V}_{zz}(u_1,u_2) = \frac{1}{T}\sum_{t=1}^{T} Re[Z_t(u_1)Z_t(u_2)^*]\varepsilon_t^2$.

Theorem 3.3.1. *Suppose Assumptions 3.3.1-3.3.5 hold, under $\mathbb{H}_0$,*

$$\sqrt{T}\hat{A}^{IV}(u) \Rightarrow \mathcal{G}^{IV}(u)$$

where $\mathcal{G}(u)$ is a zero-mean complex-valued Gaussian process with covariance kernel

$$\begin{aligned}
\mathcal{K}^{IV}(u_1,u_2) &\equiv cov\left[\mathcal{G}^{IV}(u_1),\mathcal{G}^{IV}(u_2)^*\right] \\
&= \gamma' V_{zz}(u_1,u_2)\gamma - \gamma' V_{zz}(u_1,0)\gamma[Q_{xz}\gamma]^{-1}[Q_{xz}(u_2)^*\gamma] - [\gamma' Q_{zx}(u_1)][\gamma' Q_{zx}]^{-1}\gamma' V_{zz}(0,u_2)\gamma \\
&\quad +[\gamma' Q_{zx}(u_1)][\gamma' Q_{zx}]^{-1}\gamma' V_{zz}(0,0)\gamma[Q_{xz}\gamma]^{-1}[Q_{xz}(u_2)^*\gamma],
\end{aligned}$$

where $\gamma = Q_{zz}^{-1}Q_{zx}$.

[2] In a dynamic regression model, the error terms will be serially correlated if the endogenous variables contain lagged terms of Y_t. In this chapter, we assume that the endogenous covariates do not contain lagged values of Y_t. We can weaken this restriction by using a long-run variance-covariance matrix.

The asymptotic null distribution for the DFT $\sqrt{T}\hat{A}^{IV}(u)$ is similar to Theorem 3.2.1 expect that the covariance kernels are different. When endogenous covariates are present, we need to first project X_t on Z_t. Such estimation uncertainty is represented by γ. When both X_t and Z_t are weakly stationary, we can show

$$\mathcal{K}^{IV}(u_1, u_2) = \gamma' E\left(Z_t Z_t' \varepsilon_t^2\right) \gamma \widetilde{cov}(e^{iu_1 2\pi\tau}, e^{-iu_2 2\pi\tau}).$$

The covariance kernel for the DFT $\sqrt{T}\hat{A}^{IV}(u)$ is equal to the variance of score function in the IV regression multiplied by a pseudo-covariance introduced by the Fourier transform.

Next, we give the limit of the DFT $\hat{A}^{IV}(u)$ under $\mathbb{H}_A$.

Theorem 3.3.2. *Suppose Assumption 3.3.1 to 3.3.5 and 3.2.6 hold, under $\mathbb{H}_A$,*

$$\sup_{u\in\mathbb{R}} \left\|\hat{A}^{IV}(u) - \tilde{A}^{IV}(u)\right\| \xrightarrow{p} 0,$$

where $\tilde{A}^{IV}(u) = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T}\left(I_d e^{iu2\pi t/T} - \gamma' Q_{zx}(u)[\gamma' Q_{zx}]^{-1}\right)\gamma' E(Z_t X_t') e^{iu2\pi t/T}\beta_t$ for all $u \in \mathbb{R}$, as $T \to \infty$.

Theorem 3.3.2 shows the probability limit of the DFT is a nonzero spectrum in the frequency domain, and it is a weighted average of the unknown parameter β_t. If we let X_t and Z_t be weakly stationary, and $\beta_t \equiv \beta(\frac{t}{T})$ be a smooth function of rescaled time $t/T \in (0, 1]$, then the limit of $\hat{A}^{IV}(u)$ is

$$\begin{aligned} \tilde{A}^{IV}(u) &= \gamma' E(Z_t X_t') \lim_{T\to\infty}\left[\frac{1}{T}\sum_{t=1}^{T}\beta_t e^{iu2\pi t/T} - \frac{1}{T}\sum_{t=1}^{T}\beta_t \frac{1}{T}\sum_{t=1}^{T} e^{iu2\pi t/T}\right] \\ &\equiv \gamma' E(Z_t X_t')\widetilde{cov}[\beta(\tau), e^{iu2\pi\tau}], \end{aligned}$$

where $\widetilde{cov}[\beta(\tau), e^{iu2\pi\tau}]$ is defined as in Section 3.2. As long as β_t is not constant with respect to time, the pseudo-covariance will be different from 0 for $u \neq 0$. It is crucial to ensure the consistency of our test based on $\hat{A}^{IV}(u)$. Compared to the results in Section 3.2, we see the magnitude of the pseudo-covariance is determined by $\gamma' E(X_t Z_t')$. And it is easy to show that $E(X_t X_t') - \gamma' E(Z_t Z_t')\gamma$ is positive semi-definite. Thus, for a given local alternative, the test based on $\hat{A}(u)$ is more powerful than $\hat{A}^{IV}(u)$. And this is consistent with the literature.

To examine the behavior of the DFT $\hat{A}^{IV}(u)$ at each frequency u, we consider the following KS and CM type test statistics:

$$\hat{K}^{IV} = T \sup_{u \in \mathbb{R}} \left\| \hat{A}^{IV}(u) \right\|^2 ,$$

and

$$\hat{C}^{IV} = T \int_{\mathbb{R}} \left\| \hat{A}^{IV}(u) \right\|^2 W(u) du,$$

where $W(\cdot)$ is a weighting function that satisfies Eq.(3.4).

Theorem 3.3.3. *Suppose Assumptions 3.3.1 to 3.3.5 hold, under $\mathbb{H}_0$, by the continuous mapping theorem*

$$\hat{K}^{IV} \xrightarrow{d} \sup_{u \in \mathbb{R}} \left\| \mathcal{G}^{IV}(u) \right\|^2 ,$$

and

$$\hat{C}^{IV} \xrightarrow{d} \int_{\mathbb{R}} \left\| \mathcal{G}^{IV}(u) \right\|^2 W(u) du.$$

Theorem 3.3.3 provides the asymptotic distributions of $\hat{K}^{IV}$ and $\hat{C}^{IV}$ under $\mathbb{H}_0$. Our test is tuning parameter free, but we need resampling methods to obtain critical values. Like in the OLS case, we follow Hansen (1996) and propose the following resampling procedures.

- Step(i). Use the sample $\{Y_t, X_t', Z_t'\}_{t=1}^T$ to estimate the model via the 2SLS, and compute $\hat{K}^{IV}$ and $\hat{C}^{IV}$;

- Step (ii). Generate $i.i.d.N(0,1)$ random variables $\{v_{bt}\}_{t=1}^{T}$ and compute $\hat{K}_b^{IV}$ and $\hat{C}_b^{IV}$ using $\hat{A}_b(u) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \hat{M}_t^{IV}(u)\hat{\varepsilon}_t v_{bt}$, where

$$\hat{M}_t^{IV}(u) = \hat{X}_t e^{iu2\pi t/T} - \hat{Q}_{\hat{x}x}(u)\hat{Q}_{\hat{x}\hat{x}}^{-1}\hat{X}_t,$$

and $\hat{\varepsilon}_t$ is the estimated residuals from the 2SLS;

- Step (iii). Repeat Step (ii) for a total of B times to obtain B bootstrap test statistics $\{\hat{K}_b^{IV}, \hat{C}_b^{IV}\}_{b=1}^{B}$;

- Step (iv). Compute the bootstrap p-values for $\hat{K}^{IV}$ and $\hat{C}^{IV}$ respectively, with $p_{B,T}^{K,IV} = \frac{1}{B}\sum_{b=1}^{B} \mathbf{1}(\hat{K}_b^{IV} > \hat{K}^{IV})$ and $p_{B,T}^{C,IV} = \frac{1}{B}\sum_{b=1}^{B} \mathbf{1}(\hat{C}_b^{IV} > \hat{C}^{IV})$, where $\mathbf{1}(\cdot)$ is the indicator function.

$p_{B,T}^{K,IV}$ and $p_{B,T}^{C,IV}$ are the conditional p-values for $\hat{K}^{IV}$ and $\hat{C}^{IV}$ via resampling for B times. By similar proof as in Corollary 3.2.1, we can show they are consistent for the true p-values.

Next, we state the asymptotic power of $\hat{K}^{IV}$ and $\hat{C}^{IV}$ against fixed alternatives.

Theorem 3.3.4. *Suppose Assumptions 3.3.1-3.3.5 and 3.2.6 hold. Then under $\mathbb{H}_A$, for any sequence of non-stochastic constants $\{c_T = o(T)\}$, as $T \to \infty$,*

$$\begin{aligned} P(\hat{K}^{IV} > c_T) &\to 1, \\ P(\hat{C}^{IV} > c_T) &\to 1. \end{aligned}$$

Theorem 3.3.4 shows that our test statistics $\hat{K}^{IV}$ and $\hat{C}^{IV}$ diverge to infinity at a speed of T when $\mathbb{H}_0$ is failed. Unlike Perron and Yamamoto (2015), $\hat{K}^{IV}$ and $\hat{C}^{IV}$ are consistent against both abrupt structural breaks and smooth structural changes. Compared to Chen's (2015) nonparametric consistent test, $\hat{K}^{IV}$ and $\hat{C}^{IV}$ avoid smoothed nonparametric estimation.

To investigate the local power property of $\hat{K}^{IV}$ and $\hat{C}^{IV}$, we consider $\mathbb{H}_{A1}$ defined in Section 3.2.

Theorem 3.3.5. *Suppose Assumptions 4.1-4.5 and $\mathbb{H}_{A1}$ with $\delta_T = T^{-1/2}$ hold. Then as $T \to \infty$,*

$$\begin{aligned}
\hat{K}^{IV} &\xrightarrow{d} \sup_{u\in\mathbb{R}} \left\| \xi^{IV}(u) + \mathcal{G}(u) \right\|^2, \\
\hat{C}^{IV} &\xrightarrow{d} \int_{\mathbb{R}} \left\| \xi^{IV}(u) + \mathcal{G}(u) \right\|^2 W(u)du,
\end{aligned}$$

where

$$\xi(u)^{IV} = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \left(I_d e^{iu2\pi t/T} - \gamma' Q_{zx}(u)[\gamma' Q_{zx}]^{-1} \right) \gamma' E(Z_t X_t') e^{iu2\pi t/T} \phi_t.$$

Theorem 3.3.5 implies that our tests $\hat{K}^{IV}$ and $\hat{C}^{IV}$ have nontrivial power against a class of local alternatives with rate $\delta_T = T^{-1/2}$. This rate is faster than the nonparametric rate $T^{-1/2}h^{-1/4}$ in Chen (2015) where the bandwidth $h \to 0$ as $T \to \infty$. Thus, our test is asymptotically more powerful than Chen's (2015) nonparametric test.

3.4 Simulation Studies

In this section, we show the finite sample performance of our tests and compare with existing consistent tests. We consider two cases for the simulation study:

- **Case 1**: Time series regression models with purely **exogenous** covariates X_t;
- **Case 2**: Time series regression models with **endogenous** covariates X_t.

3.4.1 Exogenous Covariates

To examine the size performance of our test statistics, we adopt the following linear regression model considered in Chen and Hong (2012)

$$\text{DGP S.1}: \; Y_t = 1 + 0.5X_t + \varepsilon_t,$$

with three types of disturbance $\{\varepsilon_t\}$:

- IID Errors:

$$\varepsilon_t \sim i.i.d.N(0,1);$$

- ARCH Errors:

$$\begin{aligned}
\varepsilon_t &= \sqrt{h_t}\nu_t, \\
h_t &= 0.2 + 0.5\varepsilon_{t-1}^2, \\
\nu_t &\sim i.i.d.N(0,1);
\end{aligned}$$

- Conditional heteroskedasticity Errors:

$$\begin{aligned}
\varepsilon_t &= \sqrt{h_t}\nu_t, \\
h_t &= 0.2 + 0.5X_t^2, \\
\nu_t &\sim i.i.d.N(0,1).
\end{aligned}$$

To show that our tests are robust to covariates with smooth structural changes and abrupt structural breaks, we consider the following three specifications for X_t:

- No structural change:

$$X_t = 0.5X_{t-1} + \nu_t;$$

- Smooth structural change:

$$X_t = 1 + \alpha(t/T)X_{t-1} + v_t,$$

where $\alpha(\tau) = 1.5 - 1.5exp[-3(\tau - 0.5)^2]$;

- Abrupt structural break:

$$X_t = \begin{cases} 1.5 + 0.5X_{t-1} + v_t, & \text{if } t \leq 0.4T, \\ -0.3X_{t-1} + v_t, & \text{otherwise.} \end{cases}$$

The innovation $v_t \sim i.i.d.N(0, 1)$ is independent of model disturbance ε_t.

To investigate the finite sample power of $\hat{K}$ and $\hat{C}$, we consider the following DGPs:

- DGP P.1—Single structural break:

$$Y_t = \begin{cases} 1.5 + 0.5X_t + \varepsilon_t, & \text{if } t \leq 0.3T, \\ 1.2 + X_t + \varepsilon_t, & \text{otherwise;} \end{cases}$$

- DGP P.2—Smooth structural change:

$$Y_t = \theta\left(\frac{t}{T}\right)(1 + 0.5X_t) + \varepsilon_t,$$

where $\theta(\tau) = 2\tau + exp[-16(\tau - 0.5)^2] - 1$;

- DGP P.3—Multiple structural breaks:

$$Y_t = \begin{cases} 0.6 + 0.3X_t + \varepsilon_t, & \text{if } 0.1T \leq t \leq 0.2T \text{ or } 0.7T \leq t \leq 0.8T \\ 1.5 + X_t + \varepsilon_t, & \text{if } 0.4T \leq t \leq 0.5T, \\ 1 + 0.5X_t + \varepsilon_t, & \text{otherwise.} \end{cases}$$

DGP P.1 to DGP P.3 cover both the abrupt structural breaks and smooth structural changes. For each DGP, we generate 1000 data sets of the random sample

$\{X_t, Y_t\}_{t=1}^T$ for $T = 100$ and 300. The computation of $\hat{K}$ involves searching over all possible $u \in \mathbb{R}$ which is computationally infeasible. Thanks to the periodicity and symmetry of the trigonometric function, we can compute $\hat{K}$ using a finite interval $\mathcal{U}$ that covers certain full periods of the sine and cosine waves. In this simulation study, we simply choose $\mathcal{U} = [0.01, 0.02, ..., 1]$. One can choose different grids based on specific problems, and it will only have a trivial impact on the result. We compute $\hat{K}$ as the following:

$$\begin{aligned}\hat{K} &= \max_{u\in\mathcal{U}} T\hat{A}(u)'\hat{A}(u)^* \\ &= \max_{u\in\mathcal{U}} \frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T} X_s'X_t\hat{\varepsilon}_s\hat{\varepsilon}_t cos\left[\frac{2\pi u(s-t)}{T}\right].\end{aligned}$$

For $\hat{C}$, we use the standard normal density weighting function $W(u) = \frac{1}{\sqrt{2\pi}}exp(\frac{u^2}{2})$ and compute it as the following:

$$\begin{aligned}\hat{C} &= T\int_{\mathbb{R}} \hat{A}(u)'\hat{A}(u)^* W(u)du \\ &= \frac{1}{T}\sum_{s=1}^{T}\sum_{t=1}^{T} X_s'X_t\hat{\varepsilon}_s\hat{\varepsilon}_t exp\left[-\frac{2\pi^2(s-t)^2}{T^2}\right].\end{aligned}$$

The critical values of $\hat{K}$ and $\hat{C}$ are obtained by the resampling method described in Section 3.2, and we set the number of resampling in each replication to be 499. To show the efficiency gain of our tests, we compare the rejection rate of our tests with consistent test $\hat{H}$ by Chen and Hong (2012). The comparison is made for the case that X_t is weakly stationary because their test is not robust to unknown structural change in the covariates. We compute $\hat{H}$ using the heteroskedasticity robust variance. We choose the bandwidth h to be $T^{-1/5}/\sqrt{12}$ by the rule of thumb, and use the uniform kernel. The critical values are computed by the resampling method (wild bootstrap) provided in Chen and Hong (2012).

Table B.1 and B.2 provide the size performance of $\hat{K}$ and $\hat{C}$ when the sample size $T = 100$ and 300. We can see $\hat{K}$ and $\hat{C}$ are robust to unknown structural change in X_t and unknown conditional volatility dynamics of ε_t. The empirical rejection rates of both $\hat{K}$ and $\hat{C}$ are close to the nominal levels in both small and large samples. Compared to the tests that require smoothed nonparametric estimation of time-varying coefficients, our tests based on the whole sample involves less noise.

Table B.3 and B.4 provide the power performance of $\hat{K}$, $\hat{C}$, and $\hat{H}$ when the sample size $T = 100$ and 300. When there exists a single structural break, $\hat{K}$ and $\hat{C}$ are both more powerful than $\hat{H}$. And the rejection rate increases as the sample size grows from 100 to 300. When there exist smooth structural changes, $\hat{C}$ is less powerful than $\hat{K}$. Because the variation in regression coefficient is very slow in DGP P.2, and the Fourier basis functions with low frequencies can capture it well. Given $\hat{K}$ searches on a grid from 0 to 1, it contains the frequency that delivers the strongest magnitude in the periodogram. Compared to $\hat{C}$ that gives extra weights to the high frequencies, $\hat{K}$ is more powerful. When multiple structural breaks are present, $\hat{K}$ and $\hat{C}$ are still more powerful than $\hat{H}$. We can also see that for each DGP P.1 to P.3, the finite sample power increases when X_t has structural changes. Intuitively, the DFT captures not only the time-varying feature of β_t but also the time series property of X_t under $\mathbb{H}_A$. As a result, it can amplify the magnitude of the periodogram at certain frequencies and thus makes our test more powerful than the case with purely stationary covariates.

3.4.2 Endogenous Covariates

We now study the finite sample performance of $\hat{K}^{IV}$ and $\hat{C}^{IV}$ using the 2SLS when endogenous variables are present. To show that $\hat{K}^{IV}$ and $\hat{C}^{IV}$ are robust to unknown structural changes in the instruments, we let Z_t exhibit smooth structural changes:

$$Z_t = 0.7cos(2\pi t/T)Z_{t-1} + \eta_t,$$

where $\eta_t \sim i.i.d.N(0, 1)$. Here Z_t follows a locally stationary AR(1) process.

To examine the size performance, we use DGP S.1 with three types of errors specified in Section 3.4.1. Note that when the errors have conditional heteroskedasticity, we should replace X_t with Z_t. To show that our test are robust to unknown structural changes in covariates, we also consider three specifications for X_t :

- No structural change:

$$X_t = 0.5X_{t-1} + \eta_t + \varepsilon_t;$$

- Smooth structural change:

$$X_t = 1 + \alpha(t/T)X_{t-1} + \eta_t + \varepsilon_t,$$

  where $\alpha(\tau) = 1.5 - 1.5exp[-3(\tau - 0.5)^2]$;

- Abrupt structural break:

$$X_t = \begin{cases} 1.5 + 0.5X_{t-1} + \eta_t + \varepsilon_t, & \text{if } t \leq 0.4T, \\ -0.3X_{t-1} + \eta_t + \varepsilon_t, & \text{otherwise.} \end{cases}$$

To examine the power performance of $\hat{K}^{IV}$ and $\hat{C}^{IV}$, we consider DGP P.1 to DGP P.3 as listed in the exogenous covariates case which cover both the abrupt structural breaks and smooth structural changes.

By using a grid $\mathcal{U} = [0.01, 0.02, ..., 1]$ for $\hat{K}^{IV}$ and the standard normal density weighting function $W(u) = \frac{1}{\sqrt{2\pi}} exp(\frac{u^2}{2})$ for $\hat{C}^{IV}$, we have

$$\begin{aligned} \hat{K}^{IV} &= \max_{u \in \mathcal{U}} T \hat{A}^{IV}(u)' \hat{A}^{IV}(u)^* \\ &= \max_{u \in \mathcal{U}} \frac{1}{T} \sum_{s=1}^{T} \sum_{t=1}^{T} \hat{X}_s' \hat{X}_t \hat{\varepsilon}_s \hat{\varepsilon}_t cos\left[\frac{2\pi u(s-t)}{T}\right], \end{aligned}$$

and

$$\begin{aligned} \hat{C}^{IV} &= T \int_{\mathbb{R}} \hat{A}^{IV}(u)' \hat{A}^{IV}(u)^* W(u) du \\ &= \frac{1}{T} \sum_{s=1}^{T} \sum_{t=1}^{T} \hat{X}_s' \hat{X}_t \hat{\varepsilon}_s \hat{\varepsilon}_t exp\left[-\frac{2\pi^2(s-t)^2}{T^2}\right], \end{aligned}$$

respectively. Compared to the case with exogenous covariates, we just replaced X_t and $\hat{\varepsilon}_t = Y_t - X_t \hat{\beta}$ with $\hat{X}_t$ and $\hat{\varepsilon}_t = Y_t - \hat{X}_t' \hat{\beta}_{2sls}$.

We compare the finite sample power of $\hat{K}^{IV}$ and $\hat{C}^{IV}$ with the consistent test $\widehat{CH}$ proposed by Chen (2015). For each of DGPs, we generate 1000 data sets of the random sample $\{X_t, Z_t, Y_t\}_{t=1}^{T}$ for $T = 100$ and 300. The critical values of $\hat{K}^{IV}$ and $\hat{C}^{IV}$ are obtained by the resampling method described in Section 3.3, and we set the number of resampling in each replication to be 499. For the nonparametric test $\widehat{CH}$, we choose the uniform kernel and set the bandwidth h at $T^{-1/5}/\sqrt{12}$ by the rule of thumb. The critical values for $\widehat{CH}$ are computed by the resampling method provided in Chen (2015).

Table B.5 and B.6 show the size performance of $\hat{K}^{IV}$ and $\hat{C}^{IV}$ for sample size $T = 100$ and 300 respectively. They show that our tests are robust to unknown structural change in both covariates and instruments. The empirical rejection rates of $\hat{K}^{IV}$ and $\hat{C}^{IV}$ are close to the nominal levels in both small and large samples.

Table B.7 and B.8 show the power performance of $\hat{K}^{IV}$, $\hat{C}^{IV}$, and $\widehat{CH}$ in finite samples. We can see $\hat{K}^{IV}$ and $\hat{C}^{IV}$ are more powerful than $\widehat{CH}$ in all DGPs. Intuitively, the nonparametric test by Chen (2015) needs nonparametric estimation for both the structural equation and the first stage reduced form. While our tests are based on the 2SLS using the whole sample. Moreover, the unknown structural change in the covariates and instruments may help to increase the power of our tests. But it is not the case for Chen's (2015) test.

3.5 Empirical Application

In this section, we provide an empirical application to show the usefulness of our tests. In macroeconomics, the New Keynesian Phillips Curve (NKPC) has drawn much attention because it shows the short-run dynamics between inflation and real economic activity. A typical hybrid NKPC can be expressed as (Gali and Gertler, 1999):

$$\pi_t = c + \beta^f E_t \pi_{t+1} + \beta^b \pi_{t-1} + \beta^x x_t + \varepsilon_t,$$

where π_t is the rate of inflation, x_t denotes the output gap or real marginal cost, $E_t \pi_{t+1}$ is the expected inflation for the period $t+1$ given the information available up to period t, and ε_t represents the unobserved shock.

The reason it is called 'hybrid' is that it delivers a combination of the traditional Phillips curve and the NKPC based on the sticky price model (Calvo, 1983). The coefficients β^f and β^b represent the relative magnitude of forward-looking and backward-looking in the short-run inflation dynamics. Determining which factor is the main driving force is an importation issue in macroeconomics because it has different economic and policy implications. If $\beta^f > \beta^b$, then the forward-looking component dominates. It implies the the inflation should be dominated by the discounted stream of expected future marginal costs. While if $\beta^f < \beta^b$, then the backward-looking prevails. And it means inflation stickiness dominates.

In fact, many empirical studies have been done on this issue and conflicting conclusions have been reached. For instance, Gali and Gertler (1999) Gali et al. (2005), and Sbordone (2002, 2005) show that the forward-looking prevails. While Fuhrer and Moore (1995), Fuhrer (1997), Rudebusch (2002), and Estrella and Fuhrer (2002, 2003) confirm the domination of backward-looking.

One possible reason for the conflicting results obtained in various empirical studies is the structural instability in the short-run inflation dynamics. And it has been documented by several studies. Zhang et al. (2008) find breaks in 1981Q1 and 2001Q1 using US quarterly data from 1968Q3 to 2005Q4. Hall et al. (2012) find breaks in 1975Q2 and 1981Q1 using US quarterly data from 1968Q3 to 2001Q4. Chen (2015) applies a consistent test based on the whole sample and concludes that the NKPC is not stable from 1968Q3 to 2012Q4. Perron and Yamamoto (2015) find breaks in 1980Q4 and 1991Q4 using US quarterly data from 1960Q1 to 1997Q4. The existence of structural change in the NKPC is confirmed by several studies. However, the following question is still open: Is 1981 to 2001 indeed a stable period?

Using the US quarterly data from 1968Q3 to 2005Q4, we investigate the stability of the NKPC by the frequency domain based tests. Following Zhang et al. (2008) and Hall et al. (2012), we consider the following specification for the NKPC:

$$\pi_t = \beta^c + \beta^f E_t \pi_{t+1} + \beta^b \pi_{t-1} + \beta^x x_t + \sum_{j=1}^{3} \beta^j \Delta\pi_{t-j} + \varepsilon_t, \tag{3.5}$$

where $\Delta\pi_t = \pi_t - \pi_{t-1}$ removes the possible serial correlation in the error term.

We use output gap as a measure for x_t as in Hall et al. (2012) and Chen (2015). One can also use real marginal labor cost for x_t as Perron and Yamamoto (2015).

The endogenous variables are expected inflation rate ($E_t\pi_{t+1}$) and the output gap (x_t). The instruments Z_t we use are variables which contain two lags of the following variables: the expected inflation rate, the output gap, the unemployment rate, the growth rate of money aggregate M2, the short-term interest rate, and the exogenous covariates in Eq.(3.5).

The data we use are quarterly US data spanning 1968Q3 to 2005Q4. We choose the data span slightly longer than that in Hall et al. (2012) for better investigation of stability in subsamples. It is shorter than Chen (2015) to avoid the turmoil period in 2008. Below are definitions of each variable and their source.

- The inflation π_t is the annualized quarterly growth rate of the GDP deflator from the Fed;
- The expected inflation rate $E_t\pi_{t+1}$ is obtained from Greenbook one-quarter ahead forecast of inflation from the Fed;
- Output gap x_t is from the Congressional Budget Office;
- The short-term interest rate is the 3-month money market rate from OECD database;
- The unemployment rate is the seasonally adjusted quarterly unemployment rate from the Fed;
- The M2 growth rate is computed by the quarterly M2 supply from the Fed.

We first test the stability of the NKPC over the whole sample following the similar steps described in Section 3.4.2. The number of bootstrapping is set at

999. The p-values based on bootstrapped critical value are 0.001 and 0.014 for $\hat{K}^{IV}$ and $\hat{C}^{IV}$. This result shows that there exists substantial evidence of structural instability of the NKPC from 1968 to 2005.

Hall et al. (2012) and Zhang et al. (2008) both identify a break in 1981. So we now split our sample to two subsamples. The first is from 1968Q3 to 1980Q4 and the second is from 1981Q1 to 2005Q4. By this design, we can investigate whether there exists any other structural instability besides the abrupt break identified in studies mentioned above. Following the same procedure as in the whole sample test, we find the p-values for the first sample is 0.013 and 0.017 using $\hat{K}^{IV}$ and $\hat{C}^{IV}$ respectively. And the p-values for the second sample is 0.012 and 0.023 using $\hat{K}^{IV}$ and $\hat{C}^{IV}$ respectively. This result shows that there also exists instability in subsamples even when the breakpoint in 1981 is identified.

Next, we consider the intercept break in 2001 documented by Zhang et al. (2008) and test the instability using the subsample spanning 1981Q1 to 2001Q4. Our tests also indicate structural instability in this period. The p-values are 0.048 and 0.052 for $\hat{K}^{IV}$ and $\hat{C}^{IV}$ respectively. Although the test by $\hat{C}^{IV}$ is not significant at 5%, we can still conclude that the stable period treated by Hall et al. (2012) and Zhang et al. (2008) should have structural instability. Furthermore, given $\hat{K}^{IV}$ is more powerful than $\hat{C}^{IV}$ in this case, we may infer that the structural change should take place in a gradual manner because we compute $\hat{K}^{IV}$ by using a grid with low frequencies. The p-value for $\hat{C}^{IV}$ is larger because it assigns more weights to higher frequencies. From this result, we conclude that it is quite possible that there exists smooth structural changes rather than abrupt structural breaks in the NKPC from 1981 to 2001. Intuitively, the micro foundations imply that the forward-looking and backward-looking coefficients are functions of each firms' pricing behavior. Even though the change may be sudden at the individual level, it is likely to behave as a smooth change at the aggregate level.

We admit that the evidence on structural change within 1981 to 2001 is not quite strong based on the p-values. Comparing to the existing literature, we show the benefit of using an asymptotically more efficient test. It reveals new things that we do not know before, but we do not want to exaggerate the result. And that does not take anything away from the big contribution of this paper.

3.6 Model Misspecification

Many existing tests for structural change are based on a correctly specified conditional model. If the model is misspecified, then the rejection of a structural change test may come from model misspecification rather than the instability of unknown parameters (Chen and Hong, 2012). Unlike the existing literature that cannot distinguish the source of rejection from structural change and model misspecification, our frequency-based approach will be free of model misspecification under certain scenarios.

Suppose the true DGP is

$$Y_t = X_t'\beta(\xi_t) + \varepsilon_t,$$

where we assume $\beta(\xi_t) : \mathbb{R}^p \to \mathbb{R}^d$ is some unknown function of random variable $\xi_t \in \mathbb{R}^p$. When $\xi_t = X_t$, then the DGP can be viewed as a nonlinear function of X_t. We assume $E(\varepsilon_t|X_t) = 0$ a.s., and ξ_t is weakly stationary.

Consider the the DFT $\hat{A}(u)$ defined in Section 3.2:

$$
\begin{aligned}
\hat{A}(u) &= \frac{1}{T}\sum_{t=1}^{T} X_t\hat{\varepsilon}_t e^{iu2\pi t/T} \\
&= \frac{1}{T}\sum_{t=1}^{T} X_t\left[X_t'\beta(\xi_t) - X_t'\hat{\beta} + \varepsilon_t\right] e^{iu2\pi t/T} \\
&= \frac{1}{T}\sum_{t=1}^{T} X_t\left[X_t'\beta(\xi_t) - X_t'\hat{Q}_{xx}^{-1}\frac{1}{T}\sum_{t=1}^{T} X_t[X_t'\beta(\xi_t) + \varepsilon_t] + \varepsilon_t\right] e^{iu2\pi t/T} \\
&= \frac{1}{T}\sum_{t=1}^{T} X_t\left[X_t'\beta(\xi_t) - X_t'\hat{Q}_{xx}^{-1}\frac{1}{T}\sum_{t=1}^{T} X_tX_t'\beta(\xi_t) - X_t'\hat{Q}_{xx}^{-1}\frac{1}{T}\sum_{t=1}^{T} X_t\varepsilon_t + \varepsilon_t\right] e^{iu2\pi t/T} \\
&= \frac{1}{T}\sum_{t=1}^{T} X_tX_t'\beta(\xi_t)e^{iu2\pi t/T} - \hat{Q}_{xx}(u)\hat{Q}_{xx}^{-1}\frac{1}{T}\sum_{t=1}^{T} X_tX_t'\beta(\xi_t) \\
&\quad +\frac{1}{T}\sum_{t=1}^{T} X_t\varepsilon_t e^{iu2\pi t/T} - \hat{Q}_{xx}(u)\hat{Q}_{xx}^{-1}\frac{1}{T}\sum_{t=1}^{T} X_t\varepsilon_t.
\end{aligned}
$$

The DFT can also be decomposed to two parts. The second component $\hat{A}_2(u)$ will converge to a zero spectrum in frequency domain due to the orthogonality condition between X_t and ε_t. And it determines the asymptotic distribution of the DFT. Under the null hypothesis of no structural change, the first component will converge to 0 in the following two scenarios:

- Scenario 1: X_t is nonstationary but is independent of ξ_t.

$$
\begin{aligned}
& \frac{1}{T}\sum_{t=1}^{T} X_tX_t'\beta(\xi_t)e^{iu2\pi t/T} - \hat{Q}_{xx}(u)\hat{Q}_{xx}^{-1}\frac{1}{T}\sum_{t=1}^{T} X_tX_t'\beta(\xi_t) \\
\xrightarrow{p}\; & Q_{xx}(u)E[\beta(\xi_t)] - Q_{xx}(u)Q_{xx}^{-1}Q_{xx}E[\beta(\xi_t)] \\
=\; & 0,
\end{aligned}
$$

  as $T \to \infty$. Intuitively, when ξ_t is weakly stationary and independent of X_t, the first component of the DFT will behave like a projection of X_t using $\hat{M}_t(u)$ defined in Section 3.2. It will converge to a zero spectrum because

$\sum_{t=1}^{T} M_t(u)X_t = 0$ for all u. Therefore, our test is robust to model misspecification when the true DGP is a functional coefficient model (Cai et al., 2000) with the dependent variable of regression coefficients being independent of X_t. In this case, the only source of rejection for our tests will be structural change.

- Scenario 2: X_t is weakly stationary.

$$
\begin{aligned}
& \frac{1}{T}\sum_{t=1}^{T} X_t X_t' \beta(\xi_t) e^{iu2\pi t/T} - \hat{Q}_{xx}(u)\hat{Q}_{xx}^{-1}\frac{1}{T}\sum_{t=1}^{T} X_t X_t' \beta(\xi_t) \\
\xrightarrow{p} \;& E[X_t X_t' \beta(\xi_t)] \int_0^1 e^{iu2\pi\tau} d\tau - Q_{xx} \int_0^1 e^{iu2\pi\tau} d\tau Q_{xx}^{-1} E[X_t X_t' \beta(\xi_t)] \\
= \;& 0,
\end{aligned}
$$

as $T \to \infty$. Intuitively, when both X_t and ξ_t are stationary, our test cannot capture any information about the unknown parameter as a function of time, so the DFT will always converge to 0 for all u.

The derivation above shows a salient feature of testing based on frequency domain analysis. The DFT we proposed in this chapter is designed particularly to detect if the unknown parameter is a deterministic function of time. Even when the model is misspecified, as long as the unknown parameter contains no information about the deterministic time trend, our test will be robust to model misspecification. In particular, when both X_t and ξ_t are weakly stationary, the only source of rejection for our test will be structural instability. Compared with the existing tests that cannot distinguish the structural change from model misspecification, the advantage of our tests is obvious.

3.7 Conclusion

This essay proposes a novel DFT-based approach to testing for structural change in a linear time series model, which is consistent against both abrupt structural breaks and smooth structural changes. It avoids smoothed nonparametric estimation of the unknown model parameter. Therefore it is tuning parameter free and can detect a class of local alternatives at the parametric rate, which is asymptotically more efficient than the existing smoothed nonparametric tests. Our approach applies to linear time series regression models with both exogenous and endogenous variables. In particular, it is robust to structural changes of unknown type in both regressors and instruments. Furthermore, we show that our tests are robust to certain model misspecification such that the only source of rejection is the structural change.

Our tests can also be extended to many other frameworks where the constancy of unknown parameters is investigated. The only input we need is sample moment condition which is identical to 0 implied by the estimation method. To extract the local information contained in the unknown parameter, we can use the DFT as a filter to capture it. The testing method we employed in the linear time series model is based on the first order condition, which is the moment condition. That implies we can test the constancy of parameter characterized by moment conditions for multiple equations (e.g., DSGE models). We can also apply our testing method to test structural change in volatility dynamics, where the only input we need is the score function. Compared to the test based on the local QMLE in Chen and Hong (2016), our DFT-based approach is much easier to implement and will be asymptotically more powerful. Moreover, based on the discussion in Section 3.6, our tests can also examine the constancy of a functional coefficient model by replacing the Fourier basis functions of time with the Fourier basis functions of a random variable. In this way, nonparametric smoothed estimation can be avoided.

CHAPTER 4

TESTING CONSTANCY OF CONDITIONAL JOINT DEPENDENCE

4.1 Basic Framework

Let $\{Y_t, Z_t, X_t\}_{t=1}^{\infty}$ be a weakly dependent and strictly stationary time series process, where $Y_t \in \mathbb{R}^p$, $Z_t \in \mathbb{R}^q$, and $X_t \in \mathbb{R}^k$. Let i be an imaginary number with $i = \sqrt{-1}$. We are interested in testing if the joint dependence between Y_t and Z_t is constant with respect to X_t. Here we compare the distance between generalized conditional covariance function

$$\sigma(u, v, x) \equiv cov(e^{iu'Y_t}, e^{iv'Z_t}|X_t = x), \tag{4.1}$$

and generalized unconditional covariance function

$$\sigma(u, v) \equiv cov(e^{iu'Y_t}, e^{iv'Z_t}), \tag{4.2}$$

for $(u, v) \in \mathbb{R}^{p+q}$. It is well-known that the generalized (un)conditional covariance function can capture the co-movement between two random variables at any (un)conditional moments. Specifically, assuming the every moment of both Y_t and Z_t exists. We can rewrite $\sigma(u, v, x)$ and $\sigma(u, v)$ using Taylor series expansion as

$$\begin{aligned} \sigma(u, v, x) &= cov\left(\sum_{m=0}^{\infty} \frac{(iu'Y_t)^m}{m!}, \sum_{l=0}^{\infty} \frac{(iv'Z_t)^l}{l!}|X_t = x\right) \\ &= \sum_{m=0}^{\infty} \sum_{l=0}^{\infty} \frac{(iu)^m (iv)^l}{m!l!} cov\left(Y_t^m, Z_t^l|X_t = x\right) \end{aligned}$$

and

$$\begin{aligned} \sigma(u, v) &= cov\left(\sum_{m=0}^{\infty} \frac{(iu'Y_t)^m}{m!}, \sum_{l=0}^{\infty} \frac{(iv'Z_t)^l}{l!}\right) \\ &= \sum_{m=0}^{\infty} \sum_{l=0}^{\infty} \frac{(iu)^m (iv)^l}{m!l!} cov\left(Y_t^m, Z_t^l\right) \end{aligned}$$

It is straightforward to see that by comparing the generalized conditional covariance function and the generalized unconditional covariance function for each $X_t = x$, we are checking if the co-movement between Y_t and Z_t is dependent on X_t at every pair of moments.

In this chapter, the hypothesis of interest is as follows:

$$\mathbb{H}_0 : \Pr[\sigma(u, v, X_t) = \sigma(u, v)] = 1 \text{ for all } (u, v) \in \mathbb{R}^{p+q}$$

against

$$\mathbb{H}_A : \Pr[\sigma(u, v, X_t) \neq \sigma(u, v)] > 0 \text{ for some } (u, v) \in \mathbb{R}^{p+q}.$$

Let

$$\begin{aligned} \phi_{yz}(u, v, x) \equiv & \quad E(e^{iu'Y_t + iv'Z_t} | X_t = x), \\ \phi_y(u, x) \equiv & \quad E(e^{iu'Y_t} | X_t = x), \\ \text{and } \phi_z(v, x) \equiv & \quad E(e^{iv'Z_t} | X_t = x), \end{aligned}$$

be the joint conditional characteristic function (CCF) for Y_t and Z_t, marginal CCF for Y_t, and marginal CCF for Z_t, given $X_t = x$ respectively. And let

$$\begin{aligned} \phi_{yz}(u, v) \equiv & \quad E(e^{iu'Y_t + iv'Z_t}), \\ \phi_y(u) \equiv & \quad E(e^{iu'Y_t}), \\ \text{and } \phi_z(v) \equiv & \quad E(e^{iv'Z_t}), \end{aligned}$$

be the joint unconditional characteristic function (UCF) for Y_t and Z_t, marginal UCF for Y_t, and marginal UCF for Z_t, respectively. Then we can rewrite Eq.(4.1) and Eq.(4.2) using the CCF's and UCF's as the following:

$$\sigma(u, v, x) = \phi_{yz}(u, v, x) - \phi_y(u, x)\phi_z(v, x), \tag{4.3}$$

and

$$\sigma(u, v) = \phi_{yz}(u, v) - \phi_y(u)\phi_z(v). \tag{4.4}$$

We want to construct a test statistic that compares the distance between generalized conditional covariance and generalized unconditional covariance using

CCF's and UCF's. In this chapter, we follow the conventional way of using the L_2-norm and propose the following test statistic

$$T = \iiint |\sigma(u, v, x) - \sigma(u, v)|^2 a(x) f(x) dW(u, v) dx, \tag{4.5}$$

where $f(x) : \mathbb{R}^k \to \mathbb{R}$ is the probability density function for X_t , $a(x) : \mathbb{R}^k \to \mathbb{R}$ is a weighting function for X_t, and $W(u, v) : \mathbb{R}^{p+q} \to \mathbb{R}$ is a weighting function for $(u, v) \in \mathbb{R}^{p+q}$. Both $a(\cdot)$ and $W(\cdot, \cdot)$ can be chosen at the practitioner's discretion.

The benefit of using the CCF and UCF has been elaborated in Su and White (2007), where they test conditional independence between Y_t and Z_t on X_t by comparing the CCF of Y_t on Z_t and X_t and the CCF of Y_t on Z_t. In this chapter, though we are facing a different problem, the test statistic here can seen as a generalization of the result in Su and White (2007) as they mentioned that their test can be generalized to testing independence between Y_t and X_t by eliminating Z_t. In that sense, they are comparing a CCF with an UCF. Since the generalized conditional covariance function is constructed by CCF's and the generalized unconditional covariance function is constructed by UCF's, our test is closely related to Su and White (2007). However, we maintain that our testing framework is different from conditional independence test investigated by Su and White (2007), Su and White (2008), Wang and Hong (2012), etc. We allow Y_t and Z_t to be conditionally dependent on each other rather than conditionally independent. Our interest is to test if their dependence strength varies with a third variable X_t.

To see why comparing generalized conditional covariance and generalized unconditional covariance can serve this purpose, let's rewrite our test statistic T

using Eq.(4.3) and Eq.(4.4) as

$$T = \iiint \left| \sum_{l,m}^{\infty} \frac{(iu)^m (iv)^l}{m!l!} \left[cov(Y_t^m, Z_t^l | X_t = x) - cov(Y_t^m, Z_t^l) \right] \right|^2 a(x) f(x) dW(u, v). \quad (4.6)$$

Then it is straightforward to see that the test statistic is comparing the conditional covariance and unconditional covariance between Y_t and Z_t at any pairs of moments. If X_t has no impact on the dependence between Y_t and Z_t, then we have $cov(Y_t^m, Z_t^l | X_t = x) - cov(Y_t^m, Z_t^l) = 0$, for all $m, l \in \mathbb{Z}$ and for all $x \in \mathbb{R}$.

4.2 Test Statistic

Suppose we have a random sample $\{X_t, Y_t, Z_t\}_{t=1}^n$. To construct the test statistic, we need to propose consistent estimators for $\sigma(u, v, x)$ and $\sigma(u, v)$ respectively. Since the former can be regarded as a regression function, we follow Chen and Hong (2010) and Wang and Hong (2012) by using local linear regression to estimate the CCF's $\phi_{yz}(u, v, x)$, $\phi_y(u, x)$, and $\phi_z(v, x)$. For $\sigma(u, v)$, we simply use its sample average since it is a constant function of X_t.

4.2.1 Estimation for the CCF

Using nonparametric estimation frees us from the model misspecification. However the convergence rate of a nonparametric estimator is lower than a parametric one. However, we regard that as a merit since the lower convergence rate makes the test statistic's limiting distribution to be invariant to the presence of estimated parameters. This means we can apply our test to samples generated from parametric estimations with the $\sqrt{n}$-convergence rate. The is the so-called asymptotic "nuisance parameter free" property.

For the nonparametric estimation for the CCF's, we follow Chen and Hong (2010) and Wang and Hong (2012) and use local linear regression since it can reduce the bias at the boundary regions (See Fan and Gijbels, 1996). We first show how to estimate the joint CCF, $\phi_{yz}(u, v, x)$.

Let $\theta \equiv (\theta_0, \theta_1^{'})' \in \mathbb{C}^{k+1}$ be a $(k+1) \times 1$ dimension complex-valued parameter vector, $h \equiv h(n)$ be the bandwidth and $K_h(\tau) = h^{-k}K(\frac{\tau}{h})$ such that $K(\cdot) : \mathbb{R}^k \to \mathbb{R}$ is a kernel function. We want to solve the following minimization problem:

$$\min_{\theta \in \mathbb{C}^{k+1}} \sum_{t=1}^{n} |e^{iu'Y_t + iv'Z_t} - \theta_0 - \theta_1^{'}(X_t - x)|^2 K_h(X_t - x), \tag{4.7}$$

where $x \in \mathbb{R}^k$, $u \in \mathbb{R}^p$ and $v \in \mathbb{R}^q$. The solution to Eq.(4.7) can be obtained as follows:

$$\hat{\theta} \equiv \hat{\theta}(u, v, x) = (X'HX)^{-1}X'HV.$$

where X is a $n \times (k+1)$ matrix with the rth row given by $[1, (X_r - x)']$, $H = \text{diag}[K_h(X_1 - x), ..., K_h(X_n - x)]$ is a diagonal matrix and $V = [e^{iu'Y_1 + iv'Z_1}, ..., e^{iu'Y_n + iv'Z_n}]'$. Here the parameter $\hat{\theta}$ is a complex-valued function of location x and parameters u and v. The joint CCF can be estimated by $\hat{\theta}_0$:

$$\hat{\theta}_0 \equiv \hat{\phi}_{yz}(u, v, x) = \sum_{t=1}^{n} \hat{H}\left(\frac{X_t - x}{h}\right) e^{iu'Y_t + iv'Z_t}, \tag{4.8}$$

where $\hat{H}(\cdot)$ is an effective kernel which is defined as

$$\hat{H}(t) = e_1^{'} S_n^{-1} [1 \ th]' K(t)/h^k. \tag{4.9}$$

Here $e_1 = (1, 0, ..., 0)'$ is a $(k+1) \times 1$ unit vector, and $S_n = X'HX$ is a $(k+1) \times (k+1)$ matrix. It has been shown by Hjellvik et al. (1998), Chen and Hong (2010), and Wang and Hong (2012) that we can write the effective kernel as:

$$\hat{H}(t) = \frac{1}{nh^k f(x)} K(t)[1 + op(1)]. \tag{4.10}$$

After obtaining the local linear estimator for the joint CCF $\phi_{yz}(u, v, x)$, we can automatically obtain the local linear estimator for the marginal CCFs $\phi_y(u, x)$ and $\phi_z(v, x)$ by simply setting $u = 0$ and $v = 0$ respectively:

$$\begin{aligned} \hat{\phi}_y(u, x) &= \hat{\phi}_{yz}(u, 0, x) = \textstyle\sum_{t=1}^{n} \hat{H}\left(\frac{X_t - x}{h}\right) e^{iu' Y_t}, \\ \hat{\phi}_z(v, x) &= \hat{\phi}_{yz}(0, v, x) = \textstyle\sum_{t=1}^{n} \hat{H}\left(\frac{X_t - x}{h}\right) e^{iv' Z_t}. \end{aligned} \tag{4.11}$$

Combining Eq.(4.8) and Eq.(4.11) we have the following nonparametric kernel estimator for the generalized conditional covariance function:

$$\hat{\sigma}(u, v, x) = \hat{\phi}_{yz}(u, v, x) - \hat{\phi}_y(u, x)\hat{\phi}_z(v, x) \tag{4.12}$$

Here we use the same bandwidth for the estimation of the joint CCF and marginal CCF's to obtain neat asymptotic results. In practice, we may choose different bandwidth to estimate $\phi_{yz}(u, v, x)$, $\phi_y(u, x)$ and $\phi_z(v, x)$ respectively. As long as the proper rate between sample size and bandwidth is chosen, the results in this chapter remain unchanged.

4.2.2 Estimation for the UCF

For the estimation of the UCF's, we directly take their corresponding sample average as consistent estimators since they are constant functions with respect to X_t. There are two benefits of doing this. First, sample mean is easy to compute without any regression analysis. The cost of doing this is that the unconditional part has no impact on the asymptotic distribution of the test statistic. This may limit the finite sample performance of our test statistic. We can regard the unconditional part as an estimation error introduced by parametric estimation when the sample is collected from regression residuals. Second, the sample average can be seen as giving equal weights to each observations. By comparing a function of x and a constant function with respect to x, we are actually comparing a locally weighted average to its globally unweighted average.

The estimators for UCF's can be written as follows:

$$\begin{aligned}\hat{\phi}_{yz}(u,v) = & \ \tfrac{1}{n}\textstyle\sum_{t=1}^{n} e^{iu'Y_t+iv'Z_t} \\ \hat{\phi}_{y}(u) = & \ \tfrac{1}{n}\textstyle\sum_{t=1}^{n} e^{iu'Y_t} \\ \hat{\phi}_{z}(v) = & \ \tfrac{1}{n}\textstyle\sum_{t=1}^{n} e^{iv'Z_t}\end{aligned}$$

and the estimator for the generalized unconditional covariance function is

$$\hat{\sigma}(u,v) = \hat{\phi}_{yz}(u,v) - \hat{\phi}_{y}(u)\hat{\phi}_{z}(v). \tag{4.13}$$

4.2.3 The Test Statistic

After presenting the estimators for the generalized conditional covariance function and generalized unconditional covariance function, we can provide the sample version of our test statistic T as the following

$$\hat{T}_n = \frac{1}{n}\sum_{t=1}^{n}\iiint |\hat{\sigma}(u,v,X_t) - \hat{\sigma}(u,v)|^2 a(X_t) dW(u,v), \tag{4.14}$$

where $\hat{\sigma}(u,v,X_t)$ and $\hat{\sigma}(u,v)$ are defined as in Eq.(4.12) and Eq.(4.13). The standardized version of $\hat{T}_n$ is $\widehat{ST}_n$, which is defined as

$$\widehat{ST}_n = \left[nh^{k/2}\hat{T}_n - B\right] / \sqrt{V}, \tag{4.15}$$

where

$$\begin{aligned}B \ = \ & h^{-k/2}\int\Bigg[\iiint[1 - |\phi_y(u,x)|^2 - |\phi_z(v,x)|^2 + |\phi_{yz}(u,v,x)|^2 \\ & -2|\sigma(u,v)|^2]dW(u,v)\Bigg]a(x)dx\int K^2(\tau)d\tau,\end{aligned}$$

and

$$\begin{aligned} V &= 2\int a^2(x)\iiiint \Lambda(u_1,u_2,v_1,v_2,x)dW(u_1,v_1)dW(u_2,v_2)dx \\ &\quad\times \int\left|\int K(\tau)K(\tau+\eta)d\tau\right|^2 d\eta, \end{aligned}$$

where B and V are asymptotic mean and variance[1] derived under $n\to\infty$. The large sample version for our test statistic is

$$\widehat{ST}_n^L = \left[nh^{k/2}\hat{T}_n - \hat{B}_n^L\right]/\sqrt{\hat{V}_n^L},$$

where

$$\begin{aligned} \hat{B}_n^L = &\quad h^{-k/2}\int[\iint[1-|\hat{\phi}_y(u,x)|^2 - |\hat{\phi}_z(v,x)|^2 + |\hat{\phi}_{yz}(u,v,x)|^2 \\ &-2|\hat{\sigma}(u,v)|^2]dW(u,v)]a(x)dx\int K^2(\tau)d\tau, \end{aligned}$$

and

$$\begin{aligned} \hat{V}_n^L = &\quad 2\int a^2(x)\iiiint \hat{\Lambda}(u_1,u_2,v_1,v_2,x)dW(u_1,v_1)dW(u_2,v_2)dx \\ &\times\int\left|\int K(\tau)K(\tau+\eta)d\tau\right|^2 d\eta. \end{aligned}$$

where we substitute the population CCFs and UCFs with its consistent estimators. To improve the performance of the test statistic in finite sample, we also provide the following finite sample version of our test statistic

$$\widehat{ST}_n^F = \left[nh^{k/2}\hat{T}_n - \hat{B}_n^F\right]/\sqrt{\hat{V}_n^F},$$

where

$$\begin{aligned} \hat{B}_n^F = &\quad h^{k/2}\sum_{t=1}^n\sum_{s=1}^n a(X_t)\hat{H}\left(\tfrac{X_s-X_t}{h}\right)^2 \\ &\times\iint|\hat{\varepsilon}_{yz}(u,v,X_s) - \hat{\phi}_z(v,X_s)\hat{\varepsilon}_y(u,X_s) - \hat{\phi}_y(u,X_s)\hat{\varepsilon}_z(v,X_s)|dW(u,v), \end{aligned}$$

and

$$\begin{aligned} \hat{V}_n^F = &\quad 2h^{k/2}\sum_{1\le r<s\le n}\left\{\sum_{t=1}^n a(X_t)\hat{H}\left(\tfrac{X_s-X_t}{h}\right)\hat{H}\left(\tfrac{X_r-X_t}{h}\right)\right. \\ &\times\iint\Big[Re\Big[(\hat{\varepsilon}_{yz}(u,v,X_s) - \hat{\phi}_z(v,X_s)\hat{\varepsilon}_y(u,X_s) - \hat{\phi}_y(u,X_s)\hat{\varepsilon}_z(v,X_s)) \\ &\times\left.(\hat{\varepsilon}_{yz}(u,v,X_r) - \hat{\phi}_z(v,X_r)\hat{\varepsilon}_y(u,X_r) - \hat{\phi}_y(u,X_r)\hat{\varepsilon}_z(v,X_r))^*\Big]\Big]dW(u,v)\right\}^2. \end{aligned}$$

[1] $\Lambda(u_1,u_2,v_1,v_2,x)$ is simply the probability limit of variance and covariance function for the terms that determines the asymptotic distribution and see Appendix for its exact expressions.

Here we let $\varepsilon_{yz}(u, v, X_s) = e^{iu'Y_s+iv'Z_s} - \phi_{yz}(u, v, X_s)$ be the generalized regression residual and $\hat{\varepsilon}_{yz}(u, v, X_s)$ be a consistent estimator with the substitution of $\phi_{yz}(u, v, X_s)$ by $\hat{\phi}_{yz}(u, v, X_s)$. We define $\varepsilon_y(u, X_s)$ and $\varepsilon_z(v, X_s)$ in the similar way.

In the next section, we derive the asymptotic distribution of our test statistic under the null hypothesis.

4.3 Asymptotic Theory

To show our standardized test statistic is well defined and follows a standard normal distribution, we need to make the following assumptions.

Assumption 4.3.1. *Let (Ω, P) be a complete probability space.*

(a) Let the stochastic vector process $\xi_t \equiv (X_t', Y_t', Z_t')'$, $t = 1, ..., n$ be a strictly stationary and absolutely regular process on $\mathbb{R}^{p+q+k}$ with β-mixing coefficients satisfying $\sum_{j=1}^{\infty} j^2\beta(j)^{\delta/(1+\delta)} < C$ for some $0 < \delta < 1$;

(b) Let $|\cdot|_k$ denote the Euclidean norm in $\mathbb{R}^k$, we assume $E(|Y_t|_p) < \infty$, $E(|Z_t|_q) < \infty$ and $E(|Y_t|_p|X_t = x) < \infty$, $E(|Z_t|_q|X_t = x) < \infty$ for all $x \in F_x$, where F_x is a compact support set of X_t on $\mathbb{R}^k$.

(c) The marginal density function $f(x)$ of X_t is positive, bounded, continuous, and twice continuously differentiable for all $x \in F_x$.

Assumption 4.3.2. *Let $\phi_{yz}(u, v, x)$, $\phi_y(u, x)$, and $\phi_z(v, x)$ be the conditional characteristic functions of $(Y_t', Z_t')'$, Y_t, and Z_t given $X_t = x$ respectively and $\phi_{yz}(u, v)$, $\phi_y(u)$, and $\phi_z(v)$ be the unconditional characteristic function of $(Y_t', Z_t')'$, Y_t, and Z_t respectively. For each $u \in \mathbb{R}^p$, $v \in \mathbb{R}^q$, $\phi_{yz}(u, v, x)$, $\phi_y(u, x)$, and $\phi_z(v, x)$ are measurable and twice continuously differentiable with respect to x.*

Assumption 4.3.3. *Let $K : \mathbb{R}^k \to \mathbb{R}^+$ be a product kernel function of some univariate kernel k, i.e., $K(u) = \prod_{j=1}^{k} k(u_j)$, where $k : \mathbb{R} \to \mathbb{R}^+$ satisfies the Lipschitz condition and is a symmetric, bounded, and twice continuously differentiable function with $\int_{\mathbb{R}} k(u)du = 1$, $\int_{\mathbb{R}} uk(u)du = 0$, and $\int_{\mathbb{R}} u^2 k(u)du = C_k < \infty$.*

Assumption 4.3.4. *(a) $W : \mathbb{R}^{p+q} \to \mathbb{R}^+$ is a nondecreasing right continuous weighting function that weighs sets symmetric about the origin equally. Also we let $\int_{\mathbb{R}^{p+q}} ||u||^4 ||v||^4 dW(u, v) < \infty$.*

(b) $a : F_x \to \mathbb{R}$ is a bounded weighting function that is continuous over F_x, where F_x is defined as in Assumption 4.3.1.

Assumption 4.3.5. *As $n \to \infty$, the bandwidth $h \equiv h(n) \to 0$ and $h = cn^{-\lambda}$ for $\frac{1}{k+4} < \lambda < \frac{2}{3k}$ and $0 < c < \infty$.*

Assumption 4.3.1 are regularity conditions on the DGP. Assumption 4.3.1(a) impose β-mixing to restrict the degree of temporal dependence. This assumption is standard in the literature (e.g., Hjellvik et al., 1998, Su and White, 2007, and Chen and Hong, 2010) and the quadratic decay rate is weaker than an exponential one. It also guarantees us to apply the central limit theorem for degenerated U statistics for weakly dependent data (e.g., Tenreiro, 1997). Assumption 4.3.1(b) is not much involved with our asymptotic result if we let Assumption 4.3.4(a) hold. It is crucial when we weaken Assumption 4.3.4(a) and apply a very special weighting function introduced by Székely et al. (2007). See section 4.4 for details. Assumption 4.3.1(c) imposes smoothness conditions on X_t and restricts it to be a continuous random variable. This condition is imposed for the convenience for deriving asymptotic results. As demonstrated by Su and White (2008) and Wang and Hong (2012), it is possible for us to extend the results to the discrete case but that is beyond the scope of this essay and we leave it for subsequent research.

Assumption 4.3.2 imposes smoothness conditions on the conditional characteristic functions. Assumption 4.3.3 provides standard regularity conditions on kernel functions and many kernel functions such as Gaussian and Epanechnikov kernels satisfy these conditions.

Assumption 4.3.4(a) imposes a general restriction to the weighting function $W(u, v)$. It guarantees the boundedness of the integration over $(u, v) \in \mathbb{R}^{p+q}$ defined in the test statistic. This makes our derivation of asymptotic results easier. However, as we mentioned above that calculating the test statistic needs numerical integration, it is extremely computationally costly for us to calculate the integral when the p and q are large. We will propose an easier way of calculation by using a special weighting function. For this special weighting function, Assumption 4.3.4(a) will be violated. However, we can show that the test statistic is still well-defined under Assumption 4.3.1(b). Assumption 4.3.4(b) admits a flexible class of weighting functions on X_t. $a(\cdot)$ can help us to specifically investigate different aspects of X_t on driving the co-movement between Y_t and Z_t. For example, we can let $a(X_t) = \mathbf{1}(|X_t| > c)$ for some pre-specified constant c. In this case, we are investigating if the conditional joint dependence between Y_t and Z_t varies with the tail part of X_t. We can also let $a(X_t) = \mathbf{1}(X_t < 0)$ to study the effect of negative part of X_t on the the dependence between Y_t and Z_t. In addition, we can choose $a(x)$ to alleviate unreliable observations. See Wang and Hong (2012) for relevant discussions.

Assumption 4.3.5 provides the same restrictions on bandwidth as in Wang and Hong (2012). We follow their treatment to abandon the optimal bandwidth: $h \propto n^{-1/(4+k)}$ which minimizes integrated mean squared errors (IMSE). There are two benefits of doing this. One is mentioned in Wang and Hong (2012) that it reduces the number of leading terms in asymptotic mean and variance and avoids estimating the Laplacian of conditional characteristic functions by letting the bias of local linear estimator converge to 0 at a faster rate. The other reason is that our test statistic involves unconditional estimators and the cross product of the conditional and unconditional estimators. Under the restriction that $\lambda > \frac{1}{k+4}$, those unconditional terms and cross product terms become higher order term and thus have no impact on the asymptotic distribution. This makes our asymptotic result neat.

Now we provide the asymptotic distribution of $\widehat{ST}_n$ under H_0.

Theorem 4.3.1. *Suppose Assumption 4.3.1(a) and (c), and Assumptions 4.3.2 - 4.3.5 hold, then* $\widehat{ST}_n \xrightarrow{d} N(0,1)$ *under* H_0 *as* $n \to \infty$.

The proof of Theorem 4.3.1 relies on the central limit theorem for degenerated U-statistic by Tenreiro (1997). Though the proof technique is quite similar to Chen and Hong (2010) and Wang and Hong (2012) and the term that determines the asymptotic distribution is the same as that in Wang and Hong (2012), the result here is slightly different. Under the framework of testing conditional independence, the generalized conditional covariance function is equal to 0 under their null hypothesis. In our framework, we allow the generalized conditional covariance function to be a non-zero constant under H_0, this is going to change the mean and variance terms and it makes the derivation more tedious.

Let's next investigate the local asymptotic power of the test statistic. The class of local alternatives that under consideration is as follows:

$$\mathbb{H}_{A1} : \sigma(u, v, x) = \sigma(u, v) + \tau_n \kappa(u, v, x)$$

where $\kappa(u, v, x)$ satisfies:

$$\zeta \equiv \iiint |\kappa(u, v, x)|^2 a(x) f(x) dW(u, v) dx < \infty.$$

It should be noted that the local asymptotic power analysis is directly related to the convergence rate of the test statistics. This makes the local asymptotic power of our test statistic to be the same as in Wang and Hong (2012).

Theorem 4.3.2. *Suppose Assumption 4.3.1(a) and (c), Assumptions 4.3.2 - 4.3.5 and* $\mathbb{H}_{A1}$ *hold with* $\tau_n = n^{-1/2} h^{-k/4}$*, then the power of the test statistic satisfies*

$$\Pr\left[\widehat{ST} \geq z_\alpha | \mathbb{H}_{A1}(\tau_n)\right] \to 1 - \Phi(z_\alpha - \zeta / \sqrt{V})$$

where $\Phi(\cdot)$ *is the* $N(0,1)$ CDF*,* z_α *is the one side critical value of* $N(0,1)$ *at significance level* α *and* V *is as defined in Eq. (4.2.3).*

4.4 Computing the Test Statistic

The computation of the test statistic involves numerical integration over $(u, v) \in \mathbb{R}^{p+q}$ if we use a general class of weighting functions $W(u, v)$. This is computationally costly especially when the dimension of $(Y_t', Z_t')'$ is large. Though Chen and Hong (2010) suggests using a nondecreasing right-continuous function with countable discontinuity points as an alternative, this will incur efficiency lost since countably many subsets of (u, v) are missed. In this section, we will show that by using a special weighting function we can evaluate numerical integrations using Euclidean norms which greatly reduced the computation burden. It can be shown that the test statistic here is invariant to the dimension of $(Y_t', Z_t')'$ and is only affected by k which is the dimension of X_t.

The special weighting function is proposed by Székely et al. (2007). In their paper, they proposed a novel test statistic called "distance covariance" to test independence between two random variable of arbitrary dimensions. Distance covariance compares the distance between the joint UCF and the product of marginal UCF using the L_2-norm. They dealt with the high dimensional integration using a special weighting function, through which the integration over u and v can be circumvented and the computation only involves a weighted sum of Euclidean norms. Thus the computation of test statistic become invariant to the dimensions p and q. Zhou (2012) extends their test statistic into the time series context.

Since our test statistic considers both generalized unconditional covariance function and generalized conditional covariance function, this chapter can be viewed as an extension to the conditional case. And we show that the computational merit holds in our framework.

First, we state the following crucial Lemma in Székely et al. (2007).

Lemma (Székely et al., 2007) If $0 < \alpha < 2$, then for all x in $\mathbb{R}^d$

$$\int \frac{1 - \cos(v' x)}{|v|_d^{d+1}} dv = c_d |x|_d,$$

where

$$c_d = \frac{\pi^{(1+d)/2}}{\Gamma((1+d)/2)}$$

$\Gamma(\cdot)$ is the complete gamma function, and $|\cdot|_d$ is the Euclidean norm in $\mathbb{R}^d$.

We define

$$\int (\cdot) dW(u, v) \equiv \int \frac{1}{c_p |u|_p^{p+1}} \frac{1}{c_q |v|_q^{q+1}} du dv \tag{4.16}$$

where $c_p = \frac{\pi^{(1+p)/2}}{\Gamma((1+p)/2)}$ and $c_q = \frac{\pi^{(1+q)/2}}{\Gamma((1+q)/2)}$, and we want to apply this weighting function to our test statistic and make use of Lemma 1 for computation. However, the derivation of asymptotic distribution involves with Assumption 4(a) and this weighting function doesn't satisfy this assumption. Also, this weighting function is unbounded at the origin and the integration over a constant is infinity. This may make some higher order terms to be infinity. Thus we have to first show that the test statistic defined using this special weighing function is well-defined. This can be shown by the following Theorem.

Theorem 4.4.1. *Under Assumption 4.3.1(b), let $W(u, v)$ be defined as in Eq. (4.16), for any fixed $h > 0$, we have $ST_n < \infty$.*

Theorem 3 shows that our test statistic is well defined using the special weighting function. By showing the asymptotic mean and variance are both bounded, we can assure that all the higher order terms that do not determine the asymptotic distribution are also bounded. This result guarantees Theorem 4.3.1 to hold using the special weighting function. Next, we provide the expressions for the standardized test statistic under this special weighting function.

Let $\hat{H}_{st} \equiv \hat{H}\left(\frac{X_s - X_t}{h}\right)$ as defined in Eq.(4.9). For any $s, r \in (1, 2, ..., n)$, denote $P_{sr} \equiv |Y_s - Y_r|_p$ and $Q_{sr} \equiv |Z_s - Z_r|_q$. Also define

$$
\begin{aligned}
A^{(1)}_{sr} &= P_{sr} - 1/n \textstyle\sum_s P_{sr} - 1/n \sum_r P_{sr} + 1/n^2 \sum_{s,r} P_{sr} \\
B^{(1)}_{sr} &= Q_{sr} - 1/n \textstyle\sum_s Q_{sr} - 1/n \sum_r Q_{sr} + 1/n^2 \sum_{s,r} Q_{sr} \\
A^{(2)}_{srt} &= P_{sr} - \textstyle\sum_s P_{sr}H_{st} - \sum_r P_{sr}H_{rt} + \sum_{s,r} P_{sr}H_{st}H_{rt} \\
B^{(2)}_{srt} &= Q_{sr} - \textstyle\sum_s Q_{sr}H_{st} - \sum_r Q_{sr}H_{rt} + \sum_{s,r} Q_{sr}H_{st}H_{rt} \\
A^{(3)}_{srt} &= P_{sr} - 1/n \textstyle\sum_s P_{sr} - \sum_r P_{sr}H_{rt} + 1/n \sum_{s,r} P_{sr}H_{rt} \\
B^{(3)}_{srt} &= Q_{sr} - 1/n \textstyle\sum_s Q_{sr} - \sum_r Q_{sr}H_{rt} + 1/n \sum_{s,r} Q_{sr}H_{rt}.
\end{aligned}
$$

Since we use local linear regression to estimate the conditional CCF, we need to provide the exact expression for $\hat{H}_{st} \equiv \hat{H}\left(\frac{X_s - X_t}{h}\right)$. By simple derivations, we can show that

$$\hat{H}_{st} = \frac{\left[\sum_{r=1}^n K_h(X_r - X_t)(X_r - X_t)^2 - \left[\sum_{r=1}^n K_h(X_r - X_t)(X_r - X_t)\right](X_s - X_t)\right] K_h(X_s - X_t)}{\sum_{r=1}^n K_h(X_r - X_t)\left[\sum_{r=1}^n K_h(X_r - X_t)(X_r - X_t)^2\right] - \left[\sum_{r=1}^n K_h(X_r - X_t)(X_r - X_t)\right]^2},$$

where the kernel function $K_h(\cdot)$ is as defined in Assumption 4.4.3. Both local linear regression and Nadaraya-Watson estimator are simply local weighted sum of the regressands. For Nadaraya-Watson the weighting function is $K_h(\frac{X_s - X_t}{h}) / \sum_{s=1}^n K_h(\frac{X_s - X_t}{h})$ while for local linear regression the weighting function is $\hat{H}_{st}$. It is easy to verify that $\sum_s \hat{H}_{st} = 1$. This is a crucial condition to show the boundedness of the test statistic under the special weighting function. The distance covariance test statistic proposed in Székely et al., (2007) can be viewed as a special case by using equal weighting function $\hat{H}_{st} = 1/n$, for all s and t.

Theorem 4.4.2. *Under Assumption 4.3.1(b), let $W(u, v)$ be defined as in Eq. (4.16), we have*

$$\widehat{ST}_n^F = \widehat{ST}_n^W \equiv \left[h^{k/2} \sum_{t=1}^n \gamma_1(X_t) a(X_t) - \hat{B}_n^W\right] / \sqrt{\hat{V}_n^W}$$

$$\hat{B}_n^F = \hat{B}_n^W \equiv h^{k/2} \sum_{t=1}^n \sum_{s=1}^n a(X_t) \hat{H}_{st}^2 \gamma_2(X_s),$$

and

$$\hat{V}_n^F = \hat{V}_n^W \equiv 2h^{k/2} \sum_{1 \le s < t \le n} \left[\sum_{t=1}^n a(X_t) \hat{H}_{st} \hat{H}_{rt} \gamma_3(X_s, X_r)\right]^2$$

where

$$\gamma_1(X_t) = \sum\nolimits_{s,r} (A_{sr}^{(1)} B_{sr}^{(1)} / n^2 + A_{sr}^{(2)} B_{sr}^{(2)} \hat{H}_{st} \hat{H}_{rt} - 2A_{sr}^{(3)} B_{sr}^{(3)} \hat{H}_{st} / n),$$

$$\begin{aligned} \gamma_2(X_s) = \quad & \sum\nolimits_{r,l} P_{rl} Q_{rl} \hat{H}_{ls} \hat{H}_{rs} - 2\sum\nolimits_r P_{sr} Q_{sr} \hat{H}_{rs} + 6\sum\nolimits_{r,l} P_{sl} Q_{sr} \hat{H}_{rs} \hat{H}_{ls} \\ & + 2\sum\nolimits_{r,l} P_{rl} Q_{rs} \hat{H}_{rs} \hat{H}_{ls} + 2\sum\nolimits_{r,l} P_{rs} Q_{rl} \hat{H}_{rs} \hat{H}_{ls} - 4\sum\nolimits_{r,l,m} P_{rl} Q_{sm} \hat{H}_{rs} \hat{H}_{ms} \hat{H}_{ls} \\ & - 4\sum\nolimits_{r,l,m} P_{sr} Q_{lm} \hat{H}_{rs} \hat{H}_{ms} H_{ls} - 4\sum\nolimits_{r,l,m} P_{rl} Q_{rm} \hat{H}_{rs} \hat{H}_{ms} \hat{H}_{ls} \\ & + 4\sum\nolimits_{r,l,m,w} P_{rl} Q_{mw} \hat{H}_{rs} \hat{H}_{ms} \hat{H}_{ls} \hat{H}_{ws}, \end{aligned}$$

and

$$
\begin{aligned}
\gamma_3(X_s, X_r) = \quad & P_{sr}Q_{sr} - \textstyle\sum_l P_{rl}Q_{rl}\hat{H}_{ls} - \sum_l P_{sl}Q_{sl}\hat{H}_{lr} + \sum_{l,m} P_{lm}Q_{lm}\hat{H}_{ls}\hat{H}_{mr} \\
& -P_{sr}(\textstyle\sum_l Q_{lr}\hat{H}_{ls} + \sum_l Q_{sl}\hat{H}_{lr} - \sum_{l,m} Q_{lm}\hat{H}_{ls}\hat{H}_{mr}) + \sum_{l,m} P_{rl}Q_{sm}\hat{H}_{ls}\hat{H}_{mr} \\
& -Q_{sr}(\textstyle\sum_l P_{lr}\hat{H}_{ls} + \sum_l P_{sl}\hat{H}_{lr} - \sum_{l,m} P_{lm}\hat{H}_{ls}\hat{H}_{mr}) + \sum_{l,m} P_{sl}Q_{rm}\hat{H}_{lr}\hat{H}_{ms} \\
& +2\textstyle\sum_{l,m} P_{rl}Q_{rm}\hat{H}_{ls}\hat{H}_{ms} + 2\sum_{l,m} P_{sl}Q_{sm}\hat{H}_{lr}\hat{H}_{mr} + \sum_{l,m} P_{sm}Q_{lm}\hat{H}_{mr}\hat{H}_{ls} \\
& +\textstyle\sum_{l,m} P_{lm}Q_{sm}\hat{H}_{ls}\hat{H}_{mr} + \sum_{l,m} P_{rl}Q_{lm}\hat{H}_{ls}\hat{H}_{mr} + \sum_{l,m} P_{lm}Q_{lr}\hat{H}_{ls}\hat{H}_{mr} \\
& -2\textstyle\sum_{l,m,w} P_{lw}Q_{mw}\hat{H}_{ls}\hat{H}_{ms}\hat{H}_{wr} - 2\sum_{l,m,w} P_{lm}Q_{lw}\hat{H}_{ls}\hat{H}_{mr}\hat{H}_{wr} \\
& -2\textstyle\sum_{l,m} Q_{lm}\hat{H}_{ls}\hat{H}_{mr}(\sum_l P_{lr}\hat{H}_{ls} + \sum_l P_{sl}\hat{H}_{lr}) \\
& -2\textstyle\sum_{l,m} P_{lm}\hat{H}_{ls}\hat{H}_{mr}(\sum_l Q_{lr}\hat{H}_{ls} + \sum_l Q_{sl}\hat{H}_{lr}) \\
& +4\textstyle\sum_{r,l,m,w} P_{rl}Q_{mw}\hat{H}_{rs}\hat{H}_{ms}\hat{H}_{ls}\hat{H}_{ws}.
\end{aligned}
$$

Theorem 4.4.2 gives us the computational form of the finite sample version[2] of our test statistic. It is clear that the integration over $(u, v) \in \mathbb{R}^{p+q}$ has been transformed to weighted sum of Euclidean norms. This greatly simplifies the computation procedure and it makes it possible for the researchers to compare Y_t and Z_t of arbitrarily high dimensions. Though the expressions for finite sample mean $\hat{B}_n^W$ and finite sample variance $\hat{V}_n^W$ look tedious, they are simply the direct application of result in Lemma (Székely et al.,2007). They are easy to compute since all the terms in γ_1, γ_2 and γ_3 can be simply obtained by matrix multiplication.

[2]We can also provide the large sample version of the test statistic using the special weighting function. However, since the sample size is usually not sufficient large in empirical studies we only consider the finite sample case.

4.5 Simulation Studies

In this section, we study the finite sample performance of our test. We consider the following DGPs.

$$
\begin{aligned}
DGP.S1: \quad & Y_t = 0.2Y_{t-1} + \varepsilon_{1,t}; \\
& Z_t = 0.1 + 0.5Y_{t-1} + 0.5Z_{t-1} + \varepsilon_{2,t}; \\
DGP.S2: \quad & Y_t = 0.4Y_{t-1} + Z_t + \varepsilon_{1,t}; \\
& Z_t = \sqrt{h_{2,t}}\varepsilon_{2,t}, h_{2,t} = 0.01 + 0.6h_{2,t-1} + 0.2Z_{t-1}^2; \\
DGP.S3: \quad & Y_t = 0.2 + 0.4X_t + \varepsilon_{1,t}; \\
& Z_t = 0.4Z_{t-1} + \varepsilon_{2,t}; \\
DGP.P1: \quad & Y_t = 0.1 + 0.5X_tZ_{t-1} + \varepsilon_{1,t}; \\
& Z_t = 0.4Z_{t-1} + \varepsilon_{2,t}; \\
DGP.P2: \quad & Y_t = \sqrt{h_{1,t}}\varepsilon_{1,t}, h_{1,t} = 0.01 + 0.5X_t^2 + 0.1Z_t^2; \\
& Z_t = 0.4Z_{t-1} + \varepsilon_{2,t}; \\
DGP.P3: \quad & Y_t = 0.4 + 0.2X_t + 0.2\ln Z_t + \varepsilon_{1,t}; \\
& Z_t = 0.4Z_{t-1} + \varepsilon_{2,t};
\end{aligned}
$$

where $\{X_t = 0.1X_{t-1} + \varepsilon_t\}$ is an $AR(1)$ process and $\{\varepsilon_t\}$, $\{\varepsilon_{1,t}\}$, and $\{\varepsilon_{2,t}\}$ are mutually independent standard normal random variables.

In $DGP.S1$ and $DGP.S2$, Y_t and Z_t are not independent both unconditionally and conditionally on X_t. This distinguishes our test from conditional independence testing framework since the conditional independence test statistic will reject the null hypothesis of conditional independence in $DGP.S1$ and $DGP.S2$. In $DGP.S3$, only Y_t is dependent on X_t while Z_t is not. So Y_t and Z_t have zero dependence both unconditionally and conditionally. For $DGP.P1$, it is equivalent to testing functional coefficient model, where we regard X_t as the regression coefficient for Z_{t-1}. $DGP.P2$ investigates the test statistic's ability to detect the dependence in *ARCH* process. *DGP.P*3, allows us to investigate the dependence between Y_t and Z_t of different nonlinear forms and it can be viewed as testing missing variables.

For each *DGP*, we consider data set of 100, 200, and 500. We investigate the rejection rate of the test statistic based on 1000 replications. The critical value is obtained from standard normal table. We compute the finite sample version of the standardized test statistic based on Theorem 4. We choose weighing function $a(X_t) = 1$ through which we incorporate the data at the tail parts of X_t. For kernel function, we follow Wang and Hong (2012), Chen and Hong (2010) and use the Gaussian kernel: $k(u) = \frac{1}{\sqrt{2\pi}} \exp(-u^2/2)$. For bandwidth, we choose the same rate as in Wang and Hong (2012), i.e., $h = cn^{-4/17}$ since our test statistics are very relevant and have the same local asymptotic power. The difference is that we choose c to be the sample standard deviation of X_t where we follow the treatment in Hong and Lee (2013). The results are reported in Table C.1 and Table C.2. From the simulation results we can see that the test statistic has very good finite sample performance.

4.6 Extensions

In this section, we show that our test is very general and can be applied to many other econometric testing problem by proper transformations. The flexibility of our testing framework relies on the following three aspects.

First is the flexibility of choosing weighting function $a(X_t)$ and we have already demonstrated it in Section 4.3. Second is the freedom of choosing conditioning fact X_t. For example, in financial contagion literature, people are interested in checking if the co-movement between stock markets is driven by economic fundamentals (Connolly and Wang, 2003). Then we can let X_t be the proxy of real economic variables and apply our test. Also, in network literature, the link that connects each node is very crucial (Billio et al., 2012). And it may be more interesting if we can identify what factors are the link comprise of. Our test can serve this purpose by choosing X_t accordingly. Furthermore, we can define X_t to be a function of time to determine if the joint dependence varies over time. Third is moment generating property of generalized (un)conditional covariance function which allows to gauge possible reasons of rejection and make relevant inference. We will provide two testing problems to demonstrate the generality of our test.

Suppose we have a time series sample $\{Y_t', Z_t'\}_{t=1}^n$ and we are interested in testing if the joint dependence between Y_t and Z_t varies over time. In this case, we let X_t be a function of time. It would be impossible for us to apply our test statistic directly if we let $X_t = t$ since it makes no sense when sample size n goes to infinity. Here we make use of the concept "local stationarity" proposed by Dalhause (1996) and we let $X_t \equiv t/n$, which is a random variable uniformly distributed on $(0, 1]$. In this case, we have actually specified the alternative hypothesis that the dependence between Y_t and Z_t is locally stationary. And the test statistic would be $T = \iiint |\sigma(u, v, t/n) - \sigma(u, v)|^2 a(t/n) f(t/n) dW(u, v) dt$. It is straightforward to see that we are comparing the generalized conditional covariance function which is obtained locally with its unconditional counterpart which is obtained globally. If the dependence strength between Y_t and Z_t doesn't vary over time, we will expect $\sigma(u, v, t/n) = \sigma(u, v)$.

Next, let's make use of the moment generating property of characteristic functions. Suppose Y_t is 1×1 random scalar and Z_t is still $q \times 1$ vector and $E(|Y_t|)$ and $E(||Z_t||)$ exist, we take differentiation with respect to (u, v) once and let $u = v = 0$, then our test becomes testing

$$H_0 : Cov(Y_t, Z_t|t/n) = Cov(Y_t, Z_t). \text{ for all } t \in (1, 2, ..., n)$$

against

$$H_a : Cov(Y_t, Z_t|t/n) \neq Cov(Y_t, Z_t). \text{ for some } t \in (1, 2, ..., n).$$

This is indeed a test of smooth structural change proposed by Chen and Hong (2012). In their paper, there is a linear regression model

$$Y_t = Z_t'\beta(t/n) + \varepsilon_t$$

and they are testing if $\beta(t/n) = \beta$. Note that

$$\begin{aligned} Cov(Y_t, Z_t|t/n) = & \quad E(Z_tY_t|t/n) - E(Z_t|t/n)E(Y_t|t/n) \\ = & \quad E(Z_t(Z_t'\beta(t/n) + \varepsilon_t)|t/n) \\ & \quad -E(Z_t|t/n)E(Z_t'\beta(t/n) + \varepsilon_t|t/n) \\ = & \quad E(Z_tZ_t')\beta(t/n) - E(Z_t)E(Z_t')\beta(t/n) \\ = & \quad Var(Z_t)\beta(t/n), \end{aligned}$$

and by similar derivation we have $Cov(Y_t, Z_t) = Var(Z_t)\beta$. Thus, comparing the conditional and unconditional covariance is indeed comparing the regression coefficient $\beta(t/n)$ and β. Therefore, our test statistic can be applicable to this problem with proper transformations.

Likewise, suppose we don't restrict X_t to be a function of time and consider the following model

$$Y_t = Z_t'\beta(X_t) + \varepsilon_t,$$

where we want to test $H_0 : \beta(X_t) = \beta$ against $H_a : \beta(X_t) \neq \beta$. Then we are actually testing functional coefficient model[3], e.g. see Cai et al. (2000). In linear regression models, the regression coefficient captures the dependence between regressors and regressands and our test statistic can have many applications.

Strict stationarity assumption is very crucial in time series analysis. Any estimation method or test that involves the whole distribution needs this assumption to hold. Suppose we have a univariate time series $\{W_t\}_{t=1}^{\infty}$. Then $\{W_t\}_{t=1}^{\infty}$ is strictly stationary if for any admissible t_1, t_2, ..., t_p, the joint distribution of $\left\{W_{t_1}, W_{t_2}, ..., W_{t_p}\right\}$ is the same as the joint distribution of $\left\{W_{t_1+r}, W_{t_2+r}, ..., W_{t_p+r}\right\}$ for all integers r. Where $\left\{W_{t_1}, W_{t_2}, ..., W_{t_p}\right\}$ and $\left\{W_{t_1+r}, W_{t_2+r}, ..., W_{t_p+r}\right\}$ are any subsets of the original time series $\{W_t\}_{t=1}^{\infty}$. Then strict stationarity guarantees that the distribution of any subsets of $\{W_t\}_{t=1}^{\infty}$ are the same regardless of at which time point the random samples are drawn.

Here, we propose a test statistic that tests strict stationarity against local stationarity and this test can be regarded as a special case of our test statistic. The null hypothesis that $\{W_t\}_{t=1}^{\infty}$ is strictly stationary while under the alternative hypothesis, $\{W_t\}_{t=1}^{\infty}$ is stationary only at the local neighborhood of each t. It is reasonable to specify the alternative hypothesis as local stationarity since most economic structural change happens smoothly. So it is quite possible for a time series to be strictly stationary in short time period but not stationary in long time horizon. When researchers have a long time series data set, it is necessary for them to test strict stationarity against local stationarity before using economic modeling or testing based on the whole distribution. For example, copula methods are widely used in finance studies (Patton, 2012) to model the dependence between financial variables. However, it is quite possible that structural break

[3]In this case, we have to assume that X_t is independent of Z_t.

happens due to instability nature of financial markets. If the nonstationarity feature is overlooked, we may draw wrong inference by applying copula method to the whole data set.

We let $Y_t \equiv \left[W_{t_1}, W_{t_2}, ..., W_{t_p}\right]' \in \mathbb{R}^p$, the test statistic is defined as follows

$$\vartheta = \iint |E(e^{iu'Y_t}|t/T) - E(e^{iu'Y_t})|^2 a(t/n) dW(u) dt,$$

where we are comparing the conditional characteristic function with the unconditional characteristic functions. If W_t is strictly stationary, the conditional characteristic function would be the equal to the unconditional characteristic function and we expect the test statistic to be 0. If W_t is locally stationary, then the conditional characteristic function would be a function of time and we will observe large values of the test statistic.

Though this proposed test statistic is slightly different from T defined in Eq. (2.8), they are constructed under the same logic. We can show that this test statistic will follow the normal distribution with different mean and variances. We leave the details for the subsequent research.

4.7 Conclusion

We provide a test to test the constancy of conditional joint dependence by comparing the conditional generalized covariance function and the unconditional generalized covariance function. By testing if the joint dependence between to random variables is dependent of a third variable of interest, our test statistic is very useful in studying financial contagion especially when we are interest in finding the factor that drives the co-movement between financial market returns. We show that the proposed test statistic follows standard normal distribution thus we don't need to simulate the critical values which makes our test statistic favorable in empirical studies. More importantly, we use a special weighting function proposed by Székely et al. (2007) and show that the numerical integration can be transformed to Euclidean norms. This is rather an appealing feather that makes our test statistic very easy compute and the computation is invariant to possible high dimensionality. We also studies the finite sample performance of our test statistic computed using the special weighting function and its performance is good. Furthermore, we show that the testing framework provided in this paper is very general and it can be applied to many other testing problems with proper transformation. And this indicates further research directions.

APPENDIX A

APPENDIX OF CHAPTER 2

A.1 Figures and Tables

Figure A.1: Plot of the real part in Example 2.1.1

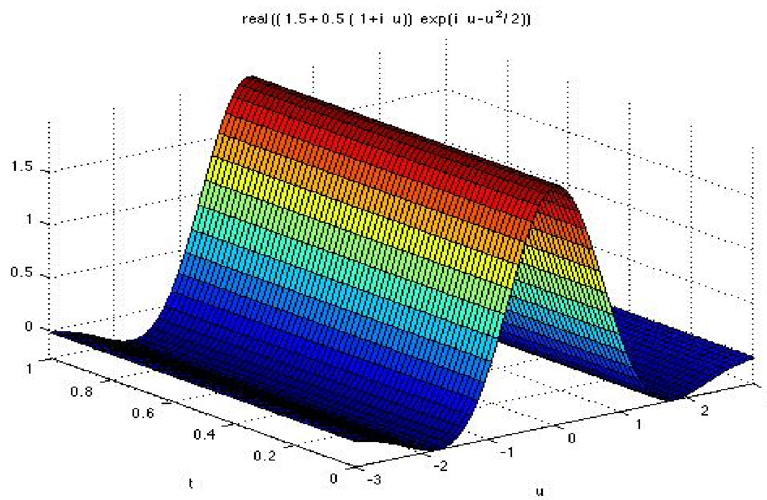

Figure A.2: Plot of the imaginary part in Example 2.1.1

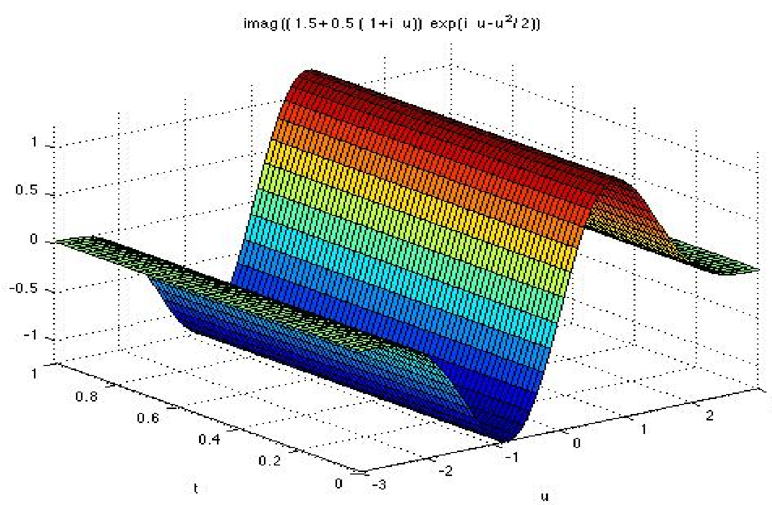

Figure A.3: Plot of the real part in Example 2.1.2

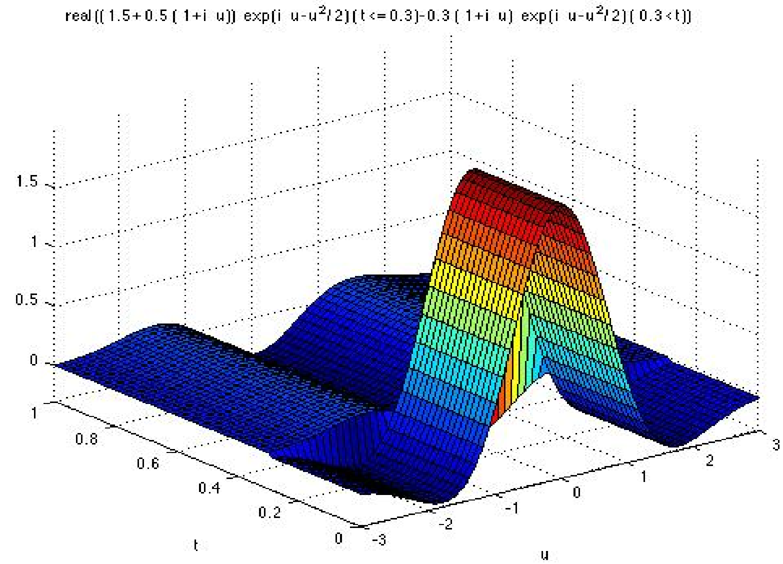

Figure A.4: Plot of the imaginary part in Example 2.1.2

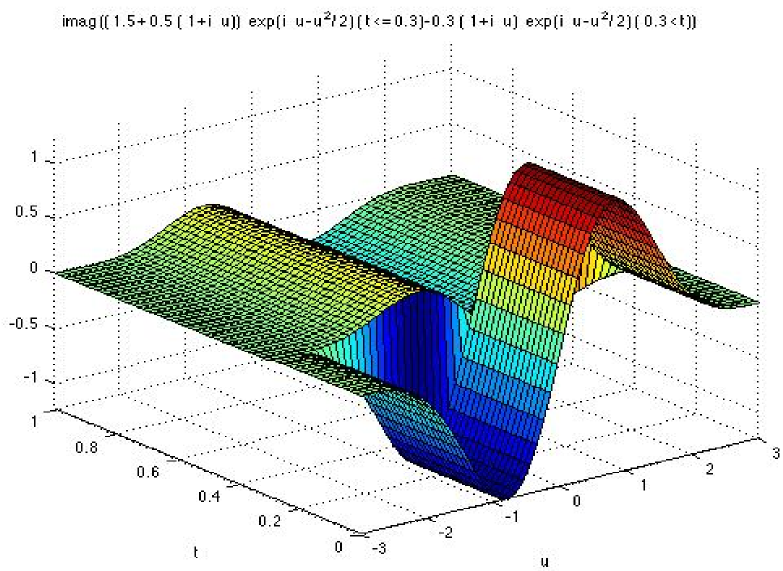

Figure A.5: Plot of the real part in Example 2.1.3

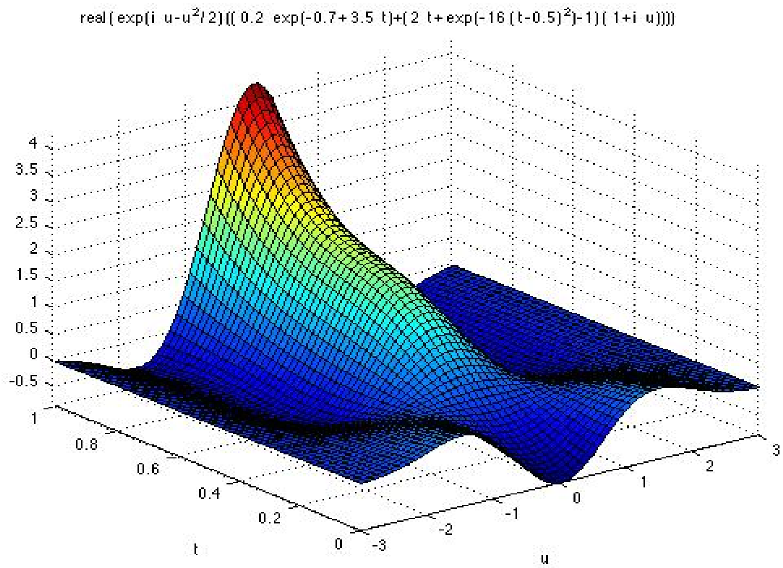

Figure A.6: Plot of the imaginary part in Example 2.1.3

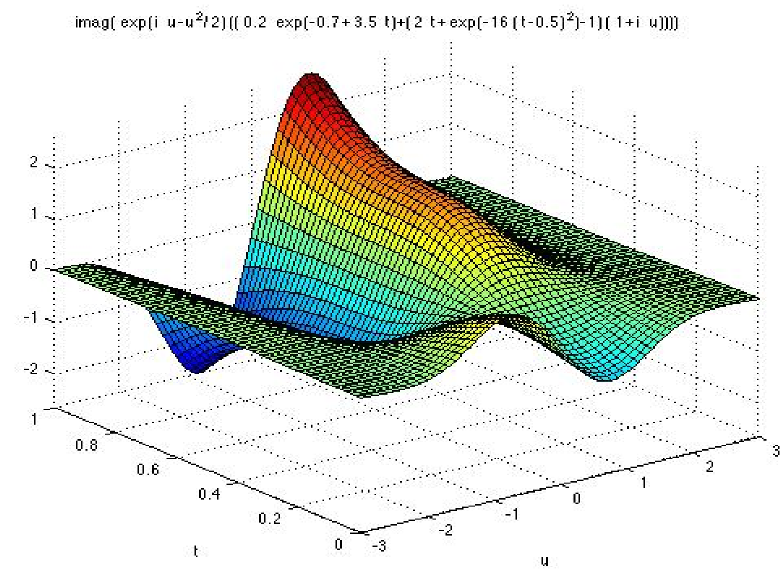

Table A.1: Empirical size in finite samples

		$\widehat{SQ}^{AS}$		$\widehat{SQ}^{BS}$		KS_T		CM_T	
		5%	10%	5%	10%	5%	10%	5%	10%
DGP S.1	T=100	0.046	0.091	0.053	0.106	0.071	0.123	0.062	0.118
	T=200	0.056	0.106	0.056	0.121	0.077	0.116	0.067	0.109
DGP S.2	T=100	0.058	0.140	0.056	0.136	0.066	0.124	0.061	0.122
	T=200	0.062	0.132	0.052	0.122	0.065	0.128	0.058	0.121
DGP S.3	T=100	0.039	0.089	0.048	0.094	0.071	0.121	0.068	0.119
	T=200	0.048	0.118	0.053	0.122	0.062	0.116	0.058	0.114
DGP S.4	T=100	0.036	0.088	0.042	0.138	——	——	——	——
	T=200	0.049	0.104	0.052	0.114	——	——	——	——
DGP S.5	T=100	0.039	0.094	0.056	0.136	——	——	——	——
	T=200	0.047	0.093	0.052	0.120	——	——	——	——

Notes: (i). Number of replication = 1000; (ii). Number of bootstrapping = 499; (iii). $\widehat{SQ}^{AS}$ and $\widehat{SQ}^{BS}$ denote the result of $\widehat{SQ}$ using asymptotic and bootstrap critical values respectively; (iv). KS_T and CM_T are the Kolmogorov-Smirnoff and Cramer-von Mises test statistics of Su and Xiao (2008); (v). The weighting function $W(u)$ is the joint $N(0, 1)$ density function; (vi). For DGP S.4 and DGP S.5, Su and Xiao's (2008) test is no longer applicable.

Table A.2: Empirical power in finite samples

		$\widehat{SQ}^{AS}$		$\widehat{SQ}^{BS}$		KS_T		CM_T	
		5%	10%	5%	10%	5%	10%	5%	10%
DGP P.1	T=100	0.588	0.794	0.512	0.768	0.818	0.886	0.710	0.836
	T=200	0.614	0.893	0.668	0.881	0.823	0.965	0.791	0.925
DGP P.2	T=100	0.900	0.968	0.776	0.932	0.972	0.990	0.944	0.982
	T=200	0.938	0.989	0.866	0.963	0.992	1.000	0.976	1.000
DGP P.3	T=100	0.226	0.448	0.316	0.530	0.116	0.168	0.086	0.168
	T=200	0.379	0.612	0.462	0.716	0.136	0.197	0.098	0.186
DGP P.4	T=100	0.248	0.394	0.138	0.318	0.166	0.288	0.174	0.294
	T=200	0.301	0.510	0.289	0.512	0.213	0.365	0.201	0.361

Notes: (i). Number of replication = 1000; (ii). Number of bootstrapping = 499; (iii). $\widehat{SQ}^{AS}$ and $\widehat{SQ}^{BS}$ denote the result of $\widehat{SQ}$ using asymptotic and bootstrap critical values respectively; (iv). KS_T and CM_T are the Kolmogorov-Smirnoff and Cramer-von Mises test statistics of Su and Xiao (2008); (v). The weighting function $W(u)$ is the joint $N(0,1)$ density function.

Table A.3: Empirical rejection rates at the 5% nominal level

		$\beta = 0.1$		$\beta = 1$		$\beta = 2$	
		$\widehat{SQ}^{AS}$	$\widehat{SQ}^{BS}$	$\widehat{SQ}^{AS}$	$\widehat{SQ}^{BS}$	$\widehat{SQ}^{AS}$	$\widehat{SQ}^{BS}$
DGPIV S.1	T=100	0.022	0.022	0.016	0.026	0.028	0.046
	T=200	0.012	0.016	0.018	0.020	0.036	0.038
DGPIV P.1	T=100	0.946	0.484	0.948	0.534	0.946	0.566
	T=200	1.000	0.590	1.000	0.650	1.000	0.612
DGPIV P.2	T=100	0.072	0.102	0.074	0.106	0.120	0.150
	T=200	0.160	0.176	0.198	0.208	0.260	0.284

Notes: (i). Number of replication = 500; (ii). Number of bootstrapping = 499; (iii). $\widehat{SQ}^{AS}$ and $\widehat{SQ}^{BS}$ denote the result of $\widehat{SQ}$ using asymptotic and bootstrap critical values respectively; (iv). The weighting function $W(u)$ is the joint $N(0,1)$ density function.

Table A.4: Empirical rejection rates at the 10% nominal level

		$\beta = 0.1$		$\beta = 1$		$\beta = 2$	
		$\widehat{SQ}^{AS}$	$\widehat{SQ}^{BS}$	$\widehat{SQ}^{AS}$	$\widehat{SQ}^{BS}$	$\widehat{SQ}^{AS}$	$\widehat{SQ}^{BS}$
DGPIV S.1	T=100	0.058	0.062	0.062	0.088	0.080	0.112
	T=200	0.048	0.056	0.054	0.072	0.096	0.114
DGPIV P.1	T=100	0.984	0.748	0.988	0.756	0.989	0.748
	T=200	1.000	0.910	1.000	0.932	1.000	0.910
DGPIV P.2	T=100	0.164	0.214	0.168	0.234	0.234	0.300
	T=200	0.274	0.302	0.316	0.360	0.396	0.468

Notes: (i). Number of replication = 500; (ii). Number of bootstrapping = 499; (iii). $\widehat{SQ}^{AS}$ and $\widehat{SQ}^{BS}$ denote the result of $\widehat{SQ}$ using asymptotic and bootstrap critical values respectively; (iv). The weighting function $W(u)$ is the joint $N(0,1)$ density function.

Table A.5: Stability test for predictive regressions

	Pre-oil-shock period			Post-oil-shock period		
	$s = 1$	$s = 6$	$s = 12$	$s = 1$	$s = 6$	$s = 12$
ds	0.008	0.0341	0.0661	0.012	0.1603	0.1543
se/p	0.001	0.0342	0.0421	0.008	0.1643	0.1563
lty	0.014	0.0281	0.0862	0.012	0.1784	0.1603
t-bill	0.016	0.0261	0.0601	0.022	0.1423	0.1784

Notes: (i). This table reports the bootstrapped p-values of $\widehat{SQ}_N$; (ii). Number of bootstrapping = 499; (iii). The predictors are default spread (ds), smoothed earning-price ratio (se/p), long-term yields on U.S. government bonds (lty), and yields on the 3-Month T-bill on the secondary market (t-bill); (iv). Pre-oil-shock period: January 1950 to December 1975; (v). Post-oil-shock period: January 1976 to December 2005; (vi).The weighting function $W(u)$ is the joint $N(0, 1)$ density function.

A.2 Mathematical Proofs

Throughout the appendix, we use the same notations as in the chapter and we denote C to be a generic bounded constant, A^* to be the complex conjugate of $A \in \mathbb{C}$, and $Re(A)$ to be the real part of $A \in \mathbb{C}$.

Proof of Lemma 2.2.1. First we show the sufficient condition: if $Pr[g_t(X_t) = g_0(X_t)] = 1$ then $\phi_t(u) = \phi_0(u)$ for all $u \in \mathbb{R}^q$.

By definition, we have $\phi_t(u) = E(Y_t e^{iu'Z_t}) = E[g_t(X_t)e^{iu'Z_t}] = E[g_0(X_t)e^{iu'Z_t}] = \phi_0(u)$, where the second to last equality is implied by the sufficient condition and the last equality comes from the definition of $\phi_0(u)$. Since this equality holds for any $u \in \mathbb{R}^q$, we thus proved the sufficient condition.

Then we show the necessary condition: only if $\phi_t(u) = \phi_0(u)$ for all $u \in \mathbb{R}^q$ then $Pr[g_t(X_t) = g_0(X_t)] = 1$.

$$
\begin{aligned}
& \phi_t(u) - \phi_0(u) \\
= & \int [g_t(x) - g_0(x)]e^{iu'z} dF_{X_t,Z_t}(x,z) \\
= & \int E[g_t(X_t) - g_0(X_t)|Z_t = z]e^{iu'z} dF_{Z_t}(z),
\end{aligned}
$$

where $F_{X_t,Z_t}(x,z)$ and $F_{Z_t}(z)$ denote the joint CDF of X_t and Z_t and marginal CDF of Z_t. Now we define $\Delta_t(z) = E[g_t(X_t) - g_0(X_t)|Z_t = z]$. We first want to show that if $\phi_t(u) = \phi_0(u)$ for all $u \in \mathbb{R}^q$, then $Pr[\Delta_t(Z_t) = 0] = 1$.

Put

$$
\delta_t^{(1)}(z) = max\{\Delta_t(z), 0\}, \qquad \delta_t^{(2)}(z) = max\{-\Delta_t(z), 0\}
$$

such that $\delta_1(z)$ and $\delta_2(z)$ are non-negative Borel measurable functions on $\mathbb{R}^q$ and

$$\Delta_t(z) = \delta_t^{(1)}(z) - \delta_t^{(2)}(z).$$

Now, let the moments

$$c_t^{(1)} = E[\delta_t^{(1)}(z)] > 0, \qquad c_t^{(2)} = E[\delta_t^{(2)}(z)] > 0.$$

Define the new probability measures ν_1 and ν_2 on the Euclidean Borel field $\mathbb{B}$ as

$$\nu_t^{(j)}(B) = \int_B \delta_t^{(j)}(z) dF_{Z_t}(z)/c_t^{(j)}, \text{ for } j = 1, 2,$$

where B is an arbitrary Borel set in $\mathbb{B}$. Now we can rewrite

$$\begin{aligned} & \phi_t(u) - \phi_0(u) \\ = & \int \Delta_t(z) e^{iu'z} dF_{Z_t}(z) \\ = & \int [\delta_t^{(1)}(z) - \delta_t^{(2)}(z)] e^{iu'z} dF_{Z_t}(z) \\ = & \int \delta_t^{(1)}(z) e^{iu'z} dF_{Z_t}(z) - \int \delta_t^{(2)}(z) e^{iu'z} dF_{Z_t}(z). \end{aligned}$$

Given $\nu_t^{(j)}(B) = \int_B \delta_t^{(j)}(z) dF_{Z_t}(z)/c_t^{(j)}$ for $j = 1, 2$, we have

$$d\nu_t^{(j)}(z) c_t^{(j)} = \delta_t^{(j)}(z) dF_{Z_t}(z).$$

Thus

$$\begin{aligned} & \phi_t(u) - \phi_0(u) \\ = & \int \delta_t^{(1)}(z) e^{iu'z} dF_{Z_t}(z) - \int \delta_t^{(2)}(z) e^{iu'z} dF_{Z_t}(z) \\ = & c_t^{(1)} \int e^{iu'z} d\nu_t^{(1)}(z) - c_t^{(2)} \int e^{iu'z} d\nu_t^{(2)}(z). \end{aligned}$$

Since $\phi_t(u) - \phi_0(u) = 0$ for all $u \in \mathbb{R}^q$ and all t, we have $c_t^{(1)} = c_t^{(2)}$ for all t by letting $u = 0$ and the fact that $\nu_t^{(1)}$ and $\nu_t^{(2)}$ are both probability measures. Therefore, we have

$$\int e^{iu'z} d\nu_t^{(1)}(z) = \int e^{iu'z} d\nu_t^{(2)}(z)$$

for all $u \in \mathbb{R}^q$ and all t. And it is equivalent to say that the two probability measure $\nu_t^{(1)}$ and $\nu_t^{(2)}$ are equal for all t. Then we have

$$\int_B \delta_t^{(1)}(z)dF_{Z_t}(z) = \int_B \delta_t^{(2)}(z)dF_{Z_t}(z),$$

for every Borel set B and all t. Furthermore, by the definition of $\delta_t^{(1)}(z)$ and $\delta_t^{(2)}(z)$, we have

$$\int_B \Delta_t(z)dF_{Z_t}(z) = 0$$

for every Borel set B and all t. Let

$$B_1 = \{z \in \mathbb{R}^q : \Delta_t(z) > 0\},$$

and

$$B_2 = \{z \in \mathbb{R}^q : \Delta_t(z) < 0\},$$

be two Borel sets. Then $B_1 \cup B_2 = \{z \in \mathbb{R}^q : \Delta_t(z) \neq 0\}$ is a null set with respect to $F_{Z_t}(z)$. Thus we have proved that if $\phi_t(u) = \phi_0(u)$ for all $u \in \mathbb{R}^q$ and all t, then $\Delta_t(z) = 0$ almost everywhere for all t. In other words, $\Delta_t(z) = 0$ is a constant function with respect to time t.

Next, we need to show that this result implies $Pr[g_t(X_t) - g_0(X_t) = 0] = 1$. Under $g_t(X_t) \neq g_0(X_t)$ for some t, we assume that the instrumental variable Z_t has to satisfy $E[g_t(X_t)|Z_t = z] = \varphi_t(z)$, where $\varphi_t(z) \equiv \varphi(z, t/T) : \mathbb{R}^q \times [0, 1] \to \mathbb{R}$ is some unknown function of both time and z. This means the time variation of $g_t(X_t)$ has to be captured by Z_t. It implies $\Delta_t(z) = E[g_t(X_t)|Z_t] - E[g_0(X_t)|Z_t]$ cannot be 0 for all t when $g_t(X_t) \neq g_0(X_t)$ for some t. Therefore, if $\Delta_t(z) = 0$ almost everywhere for all t, then $Pr[g_t(X_t) = g_0(X_t)] = 1$ and we have proved the necessary condition.

Intuitively, we have to require Z_t to be able to capture the time-varying feature of $g_t(X_t)$ via Fourier transform. Under the case of no endogeneity, this condition is satisfied automatically. $\square$

Proof of Theorem 2.3.1. We first decompose $Th^{1/2}\hat{Q}$:

$$
\begin{aligned}
Th^{1/2}\hat{Q} &= h^{1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q}\left|\hat{\phi}_t(u)-\hat{\phi}_0(u)\right|^2 W(u)du \\
&= h^{1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q}\left|\left[\hat{\phi}_t(u)-\phi_t(u)\right]-\left[\hat{\phi}_0(u)-\phi_t(u)\right]\right|^2 W(u)du \\
&= h^{1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q}\left|\hat{\phi}_t(u)-\phi_t(u)\right|^2 W(u)du + h^{1/2}\int_{\mathbb{R}^q}\sum_{t=1}^{T}\left|\hat{\phi}_0(u)-\phi_t(u)\right|^2 W(u)du \\
&\quad -2h^{1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q} Re\left\{\left[\hat{\phi}_t(u)-\phi_t(u)\right]\left[\hat{\phi}_0(u)-\phi_t(u)\right]^*\right\} W(u)du \\
&= \hat{Q}_1+\hat{Q}_2-2\hat{Q}_3.
\end{aligned}
$$

The proof of Theorem 2.3.1 consists of the proofs of Theorem A.1.1-A.1.3 below. The asymptotic distribution is determined by $\hat{Q}_1$. Under $\mathbb{H}_0$, $\hat{Q}_2$ and $\hat{Q}_3$ have no impact on the asymptotic distribution of the test statistic $Th^{1/2}\hat{Q}$. □

Theorem A.2.1. *Under the conditions of Theorem 2.3.1,* $(\hat{Q}_1 - B)/\sqrt{V} \xrightarrow{d}$ *N(0,1) as* $T \to \infty$*, where* $B = h^{-1/2}\int_{\mathbb{R}^q}|\Omega(u,u)|W(u)du\int K^2(\eta)d\eta$ *and* $V = 2\int_{\mathbb{R}^{2q}}|\Omega(u,v)|^2W(u)W(v)dudv$
$\int\left[\int K(\eta)K(\eta+\lambda)d\eta\right]^2 d\lambda$ *are asymptotic mean and variance, respectively.*

Theorem A.2.2. *Under the conditions of Theorem 2.3.1,* $\hat{Q}_2 = op(1)$.

Theorem A.2.3. *Under the conditions of Theorem 2.3.1,* $\hat{Q}_3 = op(1)$.

Proof of Theorem A.1.1. To show $(\hat{Q}_1 - B)/\sqrt{V} \xrightarrow{d}$ N(0,1) as $T \to \infty$, it suffices to show the following propositions.

Proposition A.2.1. *Under the conditions of Theorem 2.3.1,* $\hat{Q}_1 = B_1 + U + op(1)$*, where*

$$
\begin{aligned}
B_1 &= h^{-1/2}\int_{\mathbb{R}^q} E[|\varepsilon_t(u)|^2]W(u)du\int K^2(\eta)d\eta, \\
U &= \frac{1}{Th^{1/2}}\sum_{1\le s\ne r\le T}\int_{\mathbb{R}^q} Re[\varepsilon_s(u)\varepsilon_r^*(u)]W(u)du\int K(\eta)K\left(\eta+\frac{s-r}{Th}\right)d\eta.
\end{aligned}
$$

Proposition A.2.2. *Under the conditions of Theorem 2.3.1, $(U - B_2)/\sqrt{V} \xrightarrow{d}$ N(0,1) as $T \to \infty$, where*

$$
\begin{aligned}
B_2 &= 2h^{-1/2} \sum_{j=1}^{\infty} \int_{\mathbb{R}^q} E(Re[\varepsilon_t(u)\varepsilon_{t+j}^*(u)])W(u)du \int K^2(\eta)d\eta, \\
V &= 2 \int_{\mathbb{R}^{2q}} |\Omega(u,v)|^2 W(u)W(v)dudv \int \left[\int K(\eta)K(\eta+\lambda)\, d\eta \right]^2 d\lambda.
\end{aligned}
$$

Proposition A.2.3. *Under the conditions of Theorem 2.3.1, $B_1 + B_2 - \hat{B} = B - \hat{B} = op(1)$ and $V - \hat{V} = op(1)$.*

□

Proof of Proposition A.1.1. Given the equality that

$$
Y_t e^{iu'Z_t} = \phi_t(u) + \varepsilon_t(u),
$$

and $\phi_t(u) = \phi_0(u)$ under $\mathbb{H}_0$, we can rewrite $\hat{Q}_1$ as

$$
\begin{aligned}
\hat{Q}_1 &= h^{1/2} \sum_{t=1}^{T} \int_{\mathbb{R}^q} \left| \hat{\phi}_t(u) - \phi_t(u) \right|^2 W(u)du \\
&= h^{1/2} \sum_{t=1}^{T} \int_{\mathbb{R}^q} \left| \sum_{s=t-\lfloor Th \rfloor}^{t+\lfloor Th \rfloor} Y_s e^{iu'Z_s} H_{st} - \phi_t(u) \right|^2 W(u)du \\
&= h^{1/2} \sum_{t=1}^{T} \int_{\mathbb{R}^q} \left| \sum_{s=t-\lfloor Th \rfloor}^{t+\lfloor Th \rfloor} [\phi_s(u) + \varepsilon_s(u)] H_{st} - \phi_t(u) \right|^2 W(u)du \\
&= h^{1/2} \sum_{t=1}^{T} \int_{\mathbb{R}^q} \left| \sum_{s=t-\lfloor Th \rfloor}^{t+\lfloor Th \rfloor} \varepsilon_s(u) H_{st} \right|^2 W(u)du \\
&= h^{1/2} \sum_{t=1}^{T} \int_{\mathbb{R}^q} \left| \sum_{s=t-\lfloor Th \rfloor}^{t+\lfloor Th \rfloor} \frac{1}{Th} K\left(\frac{s-t}{Th}\right) \varepsilon_s(u) \right|^2 W(u)du + op(1) \\
&= h^{1/2} \sum_{t=1}^{T} \int_{\mathbb{R}^q} \left| \sum_{s=1-\lfloor Th \rfloor}^{T+\lfloor Th \rfloor} \frac{1}{Th} K\left(\frac{s-t}{Th}\right) \varepsilon_s(u) \right|^2 W(u)du + op(1) \\
&= h^{1/2} \sum_{t=1}^{T} \int_{\mathbb{R}^q} \left| \sum_{s=1}^{T} \frac{1}{Th} K\left(\frac{s-t}{Th}\right) \varepsilon_s(u) + \sum_{s=1-\lfloor Th \rfloor}^{0} \frac{1}{Th} K\left(\frac{s-t}{Th}\right) \varepsilon_s(u) \right.
\end{aligned}
$$

$$
\begin{aligned}
& +\sum_{s=T+1}^{T+\lfloor Th\rfloor}\frac{1}{Th}K\left(\frac{s-t}{Th}\right)\varepsilon_s(u)\Bigg|^2 W(u)du+op(1)\\
= \;& h^{1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q}\left|\sum_{s=1}^{T}\frac{1}{Th}K\left(\frac{s-t}{Th}\right)\varepsilon_s(u)\right|^2 W(u)du\\
& +h^{1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q}\left|\sum_{s=1-\lfloor Th\rfloor}^{0}\frac{1}{Th}K\left(\frac{s-t}{Th}\right)\varepsilon_s(u)\right|^2 W(u)du\\
& +h^{1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q}\left|\sum_{s=T+1}^{T+\lfloor Th\rfloor}\frac{1}{Th}K\left(\frac{s-t}{Th}\right)\varepsilon_s(u)\right|^2 W(u)du\\
& +2h^{1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q}\sum_{s=1}^{T}\sum_{r=1-\lfloor Th\rfloor}^{0}\frac{1}{T^2h^2}K\left(\frac{s-t}{Th}\right)K\left(\frac{r-t}{Th}\right)Re[\varepsilon_s(u)\varepsilon_r(u)^*]W(u)du\\
& +2h^{1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q}\sum_{s=1}^{T}\sum_{r=T+1}^{T+\lfloor Th\rfloor}\frac{1}{T^2h^2}K\left(\frac{s-t}{Th}\right)K\left(\frac{r-t}{Th}\right)Re[\varepsilon_s(u)\varepsilon_r(u)^*]W(u)du\\
& +2h^{1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q}\sum_{s=1-\lfloor Th\rfloor}^{0}\sum_{r=T+1}^{T+\lfloor Th\rfloor}\frac{1}{T^2h^2}K\left(\frac{s-t}{Th}\right)K\left(\frac{r-t}{Th}\right)Re[\varepsilon_s(u)\varepsilon_r(u)^*]W(u)du+op(1)\\
= \;& A_1+R_1+R_2+R_3+R_4+R_5,
\end{aligned}
$$

where we use the property that $K(\eta)=0$ for all $\eta\geq 1$ or $\eta\leq -1$. We first focus on A_1:

$$
\begin{aligned}
A_1 \;=\; & \frac{1}{T^2h^{3/2}}\sum_{t=1}^{T}\int_{\mathbb{R}^q}\left[\sum_{s=1}^{T}|\varepsilon_s(u)|^2K^2\left(\frac{s-t}{Th}\right)+\sum_{s\neq r}^{T}Re[\varepsilon_s(u)\varepsilon_r(u)^*]K\left(\frac{s-t}{Th}\right)K\left(\frac{r-t}{Th}\right)\right]W(u)du\\
= \;& \frac{1}{T^2h^{3/2}}K^2(0)\sum_{t=1}^{T}\int_{\mathbb{R}^q}|\varepsilon_t(u)|^2W(u)du+\frac{1}{T^2h^{3/2}}\sum_{s\neq t}^{T}K^2\left(\frac{s-t}{Th}\right)\int_{\mathbb{R}^q}|\varepsilon_s(u)|^2W(u)du\\
& +\frac{2}{T^2h^{3/2}}\sum_{s\neq t}^{T}K\left(\frac{s-t}{Th}\right)K(0)\int_{\mathbb{R}^q}Re[\varepsilon_s(u)\varepsilon_t(u)^*]W(u)du\\
& +\frac{1}{T^2h^{3/2}}\sum_{s\neq r\neq t}^{T}K\left(\frac{s-t}{Th}\right)K\left(\frac{r-t}{Th}\right)\int_{\mathbb{R}^q}Re[\varepsilon_s(u)\varepsilon_r(u)^*]W(u)du\\
= \;& R_6+A_2+R_7+A_3.
\end{aligned}
$$

Now we show $A_2=B_1+op(1)$:

$$
\begin{aligned}
A_2 \;=\; & \frac{1}{T^2h^{3/2}}\sum_{s\neq t}^{T}K^2\left(\frac{s-t}{Th}\right)\int_{\mathbb{R}^q}|\varepsilon_s(u)|^2W(u)du\\
= \;& \frac{2}{T^2h^{3/2}}\sum_{s=1}^{T-1}\int_{\mathbb{R}^q}|\varepsilon_s(u)|^2W(u)du\sum_{t=s+1}^{T}K^2\left(\frac{t-s}{Th}\right)
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{2}{T^2h^{3/2}}\sum_{s=1}^{T-1}\int_{\mathbb{R}^q}|\varepsilon_s(u)|^2W(u)du\sum_{j=1}^{T-s}K^2\left(\frac{j}{Th}\right)\\
&= \frac{2}{T^2h^{3/2}}\sum_{s=1}^{T-1}\int_{\mathbb{R}^q}E[|\varepsilon_s(u)|^2]W(u)du\sum_{j=1}^{T-s}K^2\left(\frac{j}{Th}\right)\\
&\quad +\frac{2}{T^2h^{3/2}}\sum_{s=1}^{T-1}\int_{\mathbb{R}^q}(|\varepsilon_s(u)|^2-E[|\varepsilon_s(u)|^2])W(u)du\sum_{j=1}^{T-s}K^2\left(\frac{j}{Th}\right)\\
&= E(A_2)+[A_2-E(A_2)]\,.
\end{aligned}
$$

For $E(A_2)$,

$$
\begin{aligned}
E(A_2) &= \frac{2}{T^2h^{3/2}}\sum_{s=1}^{T-1}\int_{\mathbb{R}^q}E[|\varepsilon_s(u)|^2]W(u)du\sum_{j=1}^{T-s}K^2\left(\frac{j}{Th}\right)\\
&= h^{-1/2}\int_{\mathbb{R}^q}E[|\varepsilon_0(u)|^2]W(u)du\times\frac{2}{Th}\sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right)K^2\left(\frac{j}{Th}\right)\\
&= h^{-1/2}\int_{\mathbb{R}^q}E[|\varepsilon_0(u)|^2]W(u)du\int K^2(\eta)d\eta+o(1),\\
&= B_1+o(1).
\end{aligned}
$$

where the last equality comes from

$$
\frac{2}{Th}\sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right)K^2\left(\frac{j}{Th}\right)=\int K^2(\eta)d\eta+o(1).
$$

For $A_2-E(A_2)$, it is straightforward to see that $E[A_2-E(A_2)]=0$. Then we just need to check $E[(A_2-E(A_2))^2]=Var(A_2)$:

$$
\begin{aligned}
Var(A_2) &= Var\left[\frac{2}{T^2h^{3/2}}\sum_{s=1}^{T-1}\int_{\mathbb{R}^q}|\varepsilon_s(u)|^2W(u)du\sum_{j=1}^{T-s}K^2\left(\frac{j}{Th}\right)\right]\\
&= \frac{4}{T^4h^3}\left[\sum_{j=1}^{T-1}(T-j)K^2\left(\frac{j}{Th}\right)\right]^2\int_{\mathbb{R}^q}\left[Var|\varepsilon_s(u)|^2+2\sum_{j=1}^{T-1}(T-j)cov\left|\varepsilon_s(u),\varepsilon_{s+j}(u)\right|\right]W(u)du\\
&\le O(T^{-2}h^{-1})+O(T^{-1}h^{-1})\\
&= o(1).
\end{aligned}
$$

where we use β-mixing conditions. Thus $A_2-E(A_2)=op(1)$ by Chebyshev's inequality.

Next we show that $A_3 = U + op(1)$:

$$
\begin{aligned}
A_3 &= \frac{1}{T^2h^{3/2}} \sum_{s\neq r\neq t}^{T} K\left(\frac{s-t}{Th}\right) K\left(\frac{r-t}{Th}\right) \int_{\mathbb{R}^q} Re[\varepsilon_s(u)\varepsilon_r(u)^*]W(u)du \\
&= \frac{1}{T^2h^{3/2}} \sum_{s\neq r\neq t}^{T} K\left(\frac{s-t}{Th}\right) K\left(\frac{s-t}{Th} + \frac{r-s}{Th}\right) \int_{\mathbb{R}^q} Re[\varepsilon_s(u)\varepsilon_r(u)^*]W(u)du \\
&= \frac{1}{Th^{1/2}} \sum_{s\neq r}^{T} \int K(\eta) K\left(\eta + \frac{s-r}{Th}\right) d\eta \int_{\mathbb{R}^q} Re[\varepsilon_s(u)\varepsilon_r(u)^*]W(u)du + op(1) \\
&= U + op(1),
\end{aligned}
$$

where the last equality comes from Riemann approximation of an integral.

At last, we have to show that $R_j = op(1)$ for $j = 1-7$.

By the symmetry of Kernel function around 0 and the fact that $K(\eta) = 0$ for all $\eta \geq 1$ or $\eta \leq -1$,

$$
\begin{aligned}
R_1 &= h^{1/2} \sum_{t=1}^{T} \int_{\mathbb{R}^q} \left| \sum_{s=1-\lfloor Th \rfloor}^{0} \frac{1}{Th} K\left(\frac{s-t}{Th}\right) \varepsilon_s(u) \right|^2 W(u)du \\
&= h^{1/2} \sum_{t=1}^{\lfloor Th \rfloor} \int_{\mathbb{R}^q} \left| \sum_{s=1-\lfloor Th \rfloor}^{0} \frac{1}{Th} K\left(\frac{s-t}{Th}\right) \varepsilon_s(u) \right|^2 W(u)du \\
&= h^{1/2} \sum_{t=1}^{\lfloor Th \rfloor} \int_{\mathbb{R}^q} \left| \sum_{s=1}^{\lfloor Th \rfloor} \frac{1}{Th} K\left(\frac{s+t-1}{Th}\right) \varepsilon_s(u) \right|^2 W(u)du \\
&= \frac{1}{T^2h^{3/2}} \sum_{s,t=1}^{\lfloor Th \rfloor} K^2\left(\frac{s+t-1}{Th}\right) \int_{\mathbb{R}^q} |\varepsilon_s(u)|^2 W(u)du \\
&\quad + \frac{1}{T^2h^{3/2}} \sum_{t=1}^{\lfloor Th \rfloor} \sum_{s\neq r}^{\lfloor Th \rfloor} K\left(\frac{s+t-1}{Th}\right) K\left(\frac{r+t-1}{Th}\right) \int_{\mathbb{R}^q} Re[\varepsilon_s(u)\varepsilon_r(u)^*]W(u)du \\
&= R_{11} + R_{12}.
\end{aligned}
$$

For R_{11}:

$$
\begin{aligned}
R_{11} &= \frac{1}{T^2h^{3/2}} \sum_{s,t=1}^{\lfloor Th \rfloor} K^2\left(\frac{s+t-1}{Th}\right) \int_{\mathbb{R}^q} |\varepsilon_s(u)|^2 W(u)du \\
&= \frac{1}{Th^{1/2}} \sum_{s=1}^{\lfloor Th \rfloor} \int_{\mathbb{R}^q} |\varepsilon_s(u)|^2 W(u)du \left[\frac{1}{Th} \sum_{t=1}^{\lfloor Th \rfloor} K^2\left(\frac{t-1}{Th} + \frac{s}{Th}\right) \right]
\end{aligned}
$$

$$= \frac{1}{Th^{1/2}} \sum_{s=1}^{\lfloor Th \rfloor} \int_{\mathbb{R}^q} |\varepsilon_s(u)|^2 W(u)du \int_0^1 K^2\left(\eta + \frac{s}{Th}\right) d\eta + o(1),$$

where the last equality comes from Riemann sum approximation of an integral.

It is straightforward to show $E(R_{11}) = o(1)$:

$$\begin{aligned}
E(R_{11}) &= \frac{1}{Th^{1/2}} \sum_{s=1}^{\lfloor Th \rfloor} \int_{\mathbb{R}^q} E[|\varepsilon_s(u)|^2] W(u)du \int_0^1 K^2\left(\eta + \frac{s}{Th}\right) d\eta + o(1) \\
&= h^{1/2} \int_{\mathbb{R}^q} E[|\varepsilon_0(u)|^2] W(u)du \left[\frac{1}{Th} \sum_{s=1}^{\lfloor Th \rfloor} \int_0^1 K^2\left(\eta + \frac{s}{Th}\right) d\eta\right] + o(1) \\
&= h^{1/2} \int_{\mathbb{R}^q} E[|\varepsilon_0(u)|^2] W(u)du \left[\int_0^1 \int_0^1 K^2(\eta + \tau)\, d\eta d\tau\right] + o(1) \\
&= O(h^{1/2}).
\end{aligned}$$

Next we show that $var(R_{11}) = o(1)$:

$$\begin{aligned}
var(R_{11}) &= var\left[\frac{1}{Th^{1/2}} \sum_{s=1}^{\lfloor Th \rfloor} \int_{\mathbb{R}^q} |\varepsilon_s(u)|^2 W(u)du \int_0^1 K^2\left(\eta + \frac{s}{Th}\right) d\eta\right] + o(1) \\
&= \frac{1}{T^2 h} \sum_{s=1}^{\lfloor Th \rfloor} var\left(\int_{\mathbb{R}^q} |\varepsilon_s(u)|^2 W(u)du\right) \left[\int_0^1 K^2\left(\eta + \frac{s}{Th}\right) d\eta\right]^2 \\
&\quad + \frac{1}{T^2 h} \sum_{s \neq r}^{\lfloor Th \rfloor} cov\left(\int_{\mathbb{R}^q} |\varepsilon_s(u)|^2 W(u)du, \int_{\mathbb{R}^q} |\varepsilon_r(u)|^2 W(u)du\right) \\
&\quad \times \left[\int_0^1 \int_0^1 K^2\left(\eta_1 + \frac{s}{Th}\right) K^2\left(\eta_2 + \frac{r}{Th}\right) d\eta_1 d\eta_2\right] + o(1) \\
&\leq \frac{1}{T^2 h} \sum_{s=1}^{\lfloor Th \rfloor} var\left(\int_{\mathbb{R}^q} |\varepsilon_s(u)|^2 W(u)du\right) \left[\int_0^1 K^2\left(\eta + \frac{s}{Th}\right) d\eta\right]^2 \\
&\quad + \frac{1}{T^2 h} \sum_{s \neq r}^{\lfloor Th \rfloor} \left|cov\left(\int_{\mathbb{R}^q} |\varepsilon_s(u)|^2 W(u)du, \int_{\mathbb{R}^q} |\varepsilon_r(u)|^2 W(u)du\right)\right| \\
&\quad \times \left[\int_0^1 \int_0^1 K^2\left(\eta_1 + \frac{s}{Th}\right) K^2\left(\eta_2 + \frac{r}{Th}\right) d\eta_1 d\eta_2\right] + o(1) \\
&\leq \frac{1}{T} var\left(\int_{\mathbb{R}^q} |\varepsilon_0(u)|^2 W(u)du\right) \frac{1}{Th} \sum_{s=1}^{\lfloor Th \rfloor} \left[\int_0^1 K^2\left(\eta + \frac{s}{Th}\right) d\eta\right]^2 \\
&\quad + \frac{2}{T^2 h} \sum_{s=1}^{\lfloor Th \rfloor - 1} \sum_{j=1}^{\lfloor Th \rfloor - s} \left|cov\left(\int_{\mathbb{R}^q} |\varepsilon_s(u)|^2 W(u)du, \int_{\mathbb{R}^q} \left|\varepsilon_{s+j}(u)\right|^2 W(u)du\right)\right| K^4(0) + o(1)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{T} var\left(\int_{\mathbb{R}^q} |\varepsilon_0(u)|^2 W(u)du\right) \int_0^1 \left[\int_0^1 K^2(\eta+\tau)\, d\eta\right]^2 d\tau \\
&\quad + \frac{K^4(0)}{T} \sum_{j=1}^{\lfloor Th \rfloor - 1} \left(1 - \frac{j}{Th}\right) \left| cov\left(\int_{\mathbb{R}^q} |\varepsilon_s(u)|^2 W(u)du, \int_{\mathbb{R}^q} \left|\varepsilon_{s+j}(u)\right|^2 W(u)du\right)\right| + o(1) \\
&\leq O(T^{-1}) + \frac{C}{T} \sum_{j=1}^{\infty} \beta(j)^{\frac{\delta}{1+\delta}} \\
&= O(T^{-1}),
\end{aligned}
$$

where the second inequality comes from the fact that $K^2(0) \geq K^2(\eta)$ for all $\eta \in [0,1]$ and the covariance sum is with respect to absolute values. The last inequality follows from the mixing inequality. By Chebyshev's inequality, we have $R_{11} = op(1)$.

For R_{12} :

$$
\begin{aligned}
R_{12} &= \frac{1}{T^2 h^{3/2}} \sum_{t=1}^{\lfloor Th \rfloor} \sum_{s \neq r}^{\lfloor Th \rfloor} K\left(\frac{s+t-1}{Th}\right) K\left(\frac{r+t-1}{Th}\right) \int_{\mathbb{R}^q} Re[\varepsilon_s(u)\varepsilon_r(u)^*] W(u)du \\
&= \frac{1}{Th^{1/2}} \sum_{s \neq r}^{\lfloor Th \rfloor} \int_{\mathbb{R}^q} Re[\varepsilon_s(u)\varepsilon_r(u)^*] W(u)du \left[\frac{1}{Th} \sum_{t=1}^{\lfloor Th \rfloor} K\left(\frac{s}{Th} + \frac{t-1}{Th}\right) K\left(\frac{r}{Th} + \frac{t-1}{Th}\right)\right] \\
&= \frac{1}{Th^{1/2}} \sum_{s \neq r}^{\lfloor Th \rfloor} \int_{\mathbb{R}^q} Re[\varepsilon_s(u)\varepsilon_r(u)^*] W(u)du \int_0^1 K\left(\frac{s}{Th} + \eta\right) K\left(\frac{r}{Th} + \eta\right) d\eta + o(1),
\end{aligned}
$$

by Riemann sum approximation of an integral. Let

$$
\begin{aligned}
\Psi_1(\Xi_s, \Xi_r) &= \int_0^1 K\left(\frac{s}{Th} + \eta\right) K\left(\frac{r}{Th} + \eta\right) d\eta \int_{\mathbb{R}^q} Re[\varepsilon_s(u)\varepsilon_r(u)^*] W(u)du \\
&\quad + \int_0^1 K\left(\frac{r}{Th} + \eta\right) K\left(\frac{s}{Th} + \eta\right) d\eta \int_{\mathbb{R}^q} Re[\varepsilon_r(u)\varepsilon_s(u)^*] W(u)du,
\end{aligned}
$$

where $\Xi_s \equiv \{Y_s, Z_s, s/T\}$, then

$$
\begin{aligned}
R_{12} &= \frac{1}{Th^{1/2}} \sum_{1 \leq s \leq r \leq \lfloor Th \rfloor} \Psi(\Xi_s, \Xi_r) \\
&= \frac{1}{Th^{1/2}} \sum_{1 \leq s \leq r \leq \lfloor Th \rfloor} \left[\Psi(\Xi_s, \Xi_r) - \Psi(\Xi_s) - \Psi(\Xi_r)\right],
\end{aligned}
$$

where $\Psi_1(\Xi_s) \equiv \int \Psi_1(\Xi_s, \Xi_r) dP(\Xi_r) = 0$, and $\Psi_1(\Xi_r) \equiv \int \Psi_1(\Xi_s, \Xi_r) dP(\Xi_s) = 0$. By Lemma A(ii) of Hjellvik et al. (1998), we have

$$\begin{aligned} & var\left[\frac{1}{Th^{1/2}} \sum_{1 \le s \le r \le \lfloor Th \rfloor} \Psi_1(\Xi_s, \Xi_r)\right] \\ \le & \frac{1}{T^2 h} T^2 h^2 E\left[|\Psi_1(\Xi_s, \Xi_r) - \Psi_1(\Xi_s) - \Psi_1(\Xi_r)|^{2(1+\delta)}\right]^{\frac{1}{1+\delta}} \sum_{j=1}^{\infty} j^2 \beta^{\frac{\delta}{1+\delta}} \\ = & o(1). \end{aligned}$$

We show $R_{12} = op(1)$ and therefore $R_1 = op(1)$.

The proof of $R_2 = op(1)$ is analogous to that of R_1. The only difference that R_2 is at the right boundary while R_1 is at the left. For space, we neglect it.

Next, we show $R_3 = op(1)$. By the property of the kernel function,

$$\begin{aligned} R_3 &= 2h^{1/2} \sum_{t=1}^{T} \int_{\mathbb{R}^q} \sum_{s=1}^{T} \sum_{r=1-\lfloor Th \rfloor}^{0} \frac{1}{T^2 h^2} K\left(\frac{s-t}{Th}\right) K\left(\frac{r-t}{Th}\right) Re[\varepsilon_s(u)\varepsilon_r(u)^*] W(u) du \\ &= \frac{2}{T^2 h^{3/2}} \sum_{t=1}^{\lfloor Th \rfloor} \left[\sum_{s=1}^{\lfloor Th \rfloor} \sum_{r=1}^{\lfloor Th \rfloor} K\left(\frac{s-t}{Th}\right) K\left(\frac{r+t-1}{Th}\right) \int_{\mathbb{R}^q} Re[\varepsilon_s(u)\varepsilon_r(u)^*] W(u) du\right] \\ &= \frac{2}{T^2 h^{3/2}} \sum_{t=1}^{\lfloor Th \rfloor} \left[\sum_{s=1}^{\lfloor Th \rfloor} K\left(\frac{s-t}{Th}\right) K\left(\frac{s+t-1}{Th}\right) \int_{\mathbb{R}^q} |\varepsilon_s(u)|^2 W(u) du\right] \\ &\quad + \frac{2}{T^2 h^{3/2}} \sum_{t=1}^{\lfloor Th \rfloor} \left[\sum_{s \ne r}^{\lfloor Th \rfloor} K\left(\frac{s-t}{Th}\right) K\left(\frac{r+t-1}{Th}\right) \int_{\mathbb{R}^q} Re[\varepsilon_s(u)\varepsilon_r(u)^*] W(u) du\right] \\ &= R_{31} + R_{32}. \end{aligned}$$

For R_{31} :

$$\begin{aligned} R_{31} &= \frac{2}{T^2 h^{3/2}} \sum_{t=1}^{\lfloor Th \rfloor} \left[\sum_{s=1}^{\lfloor Th \rfloor} K\left(\frac{s-t}{Th}\right) K\left(\frac{s+t-1}{Th}\right) \int_{\mathbb{R}^q} |\varepsilon_s(u)|^2 W(u) du\right] \\ &= \frac{1}{Th^{1/2}} \sum_{s=1}^{\lfloor Th \rfloor} \int_{\mathbb{R}^q} |\varepsilon_s(u)|^2 W(u) du \left[\frac{2}{Th} \sum_{t=1}^{\lfloor Th \rfloor} K\left(\frac{s}{Th} - \frac{t}{Th}\right) K\left(\frac{s}{Th} + \frac{t-1}{Th}\right)\right] \\ &= \frac{1}{Th^{1/2}} \sum_{s=1}^{\lfloor Th \rfloor} \int_{\mathbb{R}^q} |\varepsilon_s(u)|^2 W(u) du \int_{-1}^{1} K\left(\frac{s}{Th} - \eta\right) K\left(\frac{s}{Th} + \eta\right) d\eta + o(1). \end{aligned}$$

We first show $E(R_{31}) = o(1)$:

$$\begin{aligned}
E(R_{31}) &= \int_{\mathbb{R}^q} E[|\varepsilon_0(u)|^2]W(u)du \left[\frac{1}{Th^{1/2}} \sum_{s=1}^{\lfloor Th \rfloor} \int_{-1}^{1} K\left(\frac{s}{Th} - \eta\right) K\left(\frac{s}{Th} + \eta\right) d\eta\right] + o(1) \\
&= h^{1/2} \int_{\mathbb{R}^q} E[|\varepsilon_0(u)|^2]W(u)du \int_0^1 \int_{-1}^{1} K(\eta_1 - \eta_2) K(\eta_1 + \eta_2) d\eta_1 d\eta_2 + o(1) \\
&= O(h^{1/2}).
\end{aligned}$$

Next,

$$\begin{aligned}
var(R_{31}) &= var\left[\frac{1}{Th^{1/2}} \sum_{s=1}^{\lfloor Th \rfloor} \int_{\mathbb{R}^q} |\varepsilon_s(u)|^2 W(u)du \int_{-1}^{1} K\left(\frac{s}{Th} - \eta\right) K\left(\frac{s}{Th} + \eta\right) d\eta\right] + o(1) \\
&= \frac{1}{T^2 h} \sum_{s=1}^{\lfloor Th \rfloor} var\left[\int_{\mathbb{R}^q} |\varepsilon_s(u)|^2 W(u)du\right]\left[\int_{-1}^{1} K\left(\frac{s}{Th} - \eta\right) K\left(\frac{s}{Th} + \eta\right) d\eta\right]^2 \\
&\quad + \frac{1}{T^2 h} \sum_{s \neq r}^{\lfloor Th \rfloor} cov\left(\int_{\mathbb{R}^q} |\varepsilon_s(u)|^2 W(u)du, \int_{\mathbb{R}^q} |\varepsilon_r(u)|^2 W(u)du\right) \\
&\quad \times \left[\int_{-1}^{1} K\left(\frac{s}{Th} - \eta\right) K^2\left(\frac{s}{Th} + \eta\right) d\eta \int_{-1}^{1} K\left(\frac{r}{Th} - \eta\right) K^2\left(\frac{r}{Th} + \eta\right) d\eta\right] + o(1) \\
&= O(T^{-1}),
\end{aligned}$$

by analogous logics as in the proof of $var(R_{11}) = O(T^{-1})$. By Chebyshev's inequality, we have $R_{31} = op(1)$.

For R_{32}:

$$\begin{aligned}
R_{32} &= \frac{2}{T^2 h^{3/2}} \sum_{t=1}^{\lfloor Th \rfloor} \left[\sum_{s \neq r}^{\lfloor Th \rfloor} K\left(\frac{s-t}{Th}\right) K\left(\frac{r+t-1}{Th}\right) \int_{\mathbb{R}^q} Re[\varepsilon_s(u)\varepsilon_r(u)^*]W(u)du\right] \\
&= \frac{1}{Th^{1/2}} \sum_{s \neq r}^{\lfloor Th \rfloor} \int_{\mathbb{R}^q} Re[\varepsilon_s(u)\varepsilon_r(u)^*]W(u)du \int_{-1}^{1} K\left(\frac{s}{Th} - \eta\right) K\left(\frac{r}{Th} + \eta\right) d\eta.
\end{aligned}$$

Using analogous logics as in the proof of $R_{12} = op(1)$, we can show $R_{32} = op(1)$. Therefore, $R_3 = op(1)$. The proof of $R_4 = op(1)$ is quite analogous and we neglect it.

Next, we claim that $R_5 = 0$. The result follows directly from the fact that $K(\eta) = 0$ for all $\eta \geq 1$ or $\eta \leq -1$. Intuitively, it is a product of two kernel functions that are at left and right boundary respectively. As sample size grows to infinity, there will be no overlapping in between.

Finally, we have to show $R_6 = R_7 = op(1)$.

$$\begin{aligned} E(R_6) &= \frac{1}{T^2h^{3/2}}K^2(0)\sum_{t=1}^{T}\int_{\mathbb{R}^q} E[|\varepsilon_t(u)|^2]W(u)du \\ &= O(T^{-1}h^{3/2}) = o(1), \end{aligned}$$

and

$$\begin{aligned} var(R_6) &= \frac{1}{T^4h^3}K^4(0)\sum_{t=1}^{T} var\left[\int_{\mathbb{R}^q}|\varepsilon_t(u)|^2W(u)du\right] \\ &\quad +\frac{1}{T^4h^3}K^4(0)\sum_{s\neq t}^{T} cov\left(\int_{\mathbb{R}^q}|\varepsilon_s(u)|^2W(u)du, \int_{\mathbb{R}^q}|\varepsilon_t(u)|^2W(u)du\right) \\ &= \frac{1}{T^3h^3}K^4(0)var\left[\int_{\mathbb{R}^q}|\varepsilon_0(u)|^2W(u)du\right] \\ &\quad +\frac{1}{T^3h^3}K^4(0)\sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right)cov\left(\int_{\mathbb{R}^q}|\varepsilon_s(u)|^2W(u)du, \int_{\mathbb{R}^q}|\varepsilon_{s+j}(u)|^2W(u)du\right) \\ &\leq O(T^{-3}h^{-3}) + \frac{C}{T^3h^3}\sum_{j=1}^{\infty}\beta(j)^{\frac{\delta}{1+\delta}} \\ &= O(T^{-3}h^{-3}) = o(1). \end{aligned}$$

By Chebyshev's Inequality, $R_6 = op(1)$. $R_7 = op(1)$ follows from similar arguments as in the proof of $R_{12} = op(1)$. For space, we neglect it. □

Proof of Proposition A.1.2. Given that the U-statistic we have here exhibits possible strong dependence with nearby observations, the conventional asymptotic theory may not work out. Thus, we have to first remove the terms with possible strong dependence and consider the asymptotic behavior of the remaining term. Following Hong et al. (2014), we introduce a new tuning parameter p_T that satisfies $p_T \to \infty$, $p_T/Th \to 0$ as $T \to \infty$, and $\sum_{j=p_T}^{\infty} j^2\beta(j) \leq Cp_T^{-1}$. Denote

$\mathbb{D}_1 = \{(s,r) : 1 \leq |s-r| \leq p_T, 1 \leq s \neq r \leq T\}$ and $\mathbb{D}_2 = \{(s,r) : p_T < |s-r| < T, 1 \leq s \neq r \leq T\}$. Then we can decompose U as

$$\begin{aligned} U &= \frac{1}{Th^{1/2}} \sum_{s,r \in \mathbb{D}_1} \int K(\eta) K\left(\eta + \frac{s-r}{Th}\right) d\eta \int_{\mathbb{R}^q} Re[\varepsilon_s(u)\varepsilon_r(u)^*] W(u) du \\ &\quad + \frac{1}{Th^{1/2}} \sum_{s,r \in \mathbb{D}_2} \int K(\eta) K\left(\eta + \frac{s-r}{Th}\right) d\eta \int_{\mathbb{R}^q} Re[\varepsilon_s(u)\varepsilon_r(u)^*] W(u) du \\ &= U_1 + U_2. \end{aligned}$$

First, we show that $U_1 = B_2 + op(1)$:

$$\begin{aligned} U_1 &= \frac{1}{Th^{1/2}} \sum_{s,r \in \mathbb{D}_1} \int K^2(\eta) d\eta \int_{\mathbb{R}^q} Re[\varepsilon_s(u)\varepsilon_r(u)^*] W(u) du \\ &\quad + \frac{1}{Th^{1/2}} \sum_{s,r \in \mathbb{D}_1} \int K(\eta) \left[K\left(\eta + \frac{s-r}{Th}\right) - K(\eta)\right] d\eta \int_{\mathbb{R}^q} Re[\varepsilon_s(u)\varepsilon_r(u)^*] W(u) du \\ &= U_{11} + U_{12}. \end{aligned}$$

Now, we claim that $U_{11} = B_2 + op(1)$ and $U_{12} = op(1)$. For U_{11}, we can decompose it as $U_{11} = E(U_{11}) + U_{11} - E(U_{11})$. Next we show that $E(U_{11}) = B_2 + op(1)$ and $U_{11} - E(U_{11}) = op(1)$.

$$\begin{aligned} E(U_{11}) &= E\left\{\frac{1}{Th^{1/2}} \sum_{s,r \in \mathbb{D}_1} \int K^2(\eta) d\eta \int_{\mathbb{R}^q} Re[\varepsilon_s(u)\varepsilon_r(u)^*] W(u) du\right\} \\ &= \frac{1}{Th^{1/2}} \sum_{s=1}^{T-1} \sum_{r=s+1}^{\min\{s+p_T, T\}} \int_{\mathbb{R}^q} E\left(Re[\varepsilon_s(u)\varepsilon_r(u)^* + \varepsilon_r(u)\varepsilon_s(u)^*]\right) W(u) du \int K^2(\eta) d\eta + op(1) \\ &= h^{-1/2} \sum_{j=1}^{\min\{p_T, T-s\}} \int_{\mathbb{R}^q} \frac{1}{T} \sum_{s=1}^{T-1} E\left(Re[\varepsilon_s(u)\varepsilon_{s+j}(u)^* + \varepsilon_{s+j}(u)\varepsilon_s(u)^*]\right) W(u) du \int K^2(\eta) d\eta + op(1) \\ &= h^{-1/2} \sum_{j=1}^{\min\{p_T, T-s\}} \int_{\mathbb{R}^q} E\left(Re[\varepsilon_0(u)\varepsilon_j(u)^* + \varepsilon_j(u)\varepsilon_0(u)^*]\right) W(u) du \int K^2(\eta) d\eta + op(1) \\ &= 2h^{-1/2} \sum_{j=1}^{\infty} \int_{\mathbb{R}^q} E\left(Re[\varepsilon_0(u)\varepsilon_j(u)^*]\right) W(u) du \int K^2(\eta) d\eta + op(1) \\ &= B_2 + op(1). \end{aligned}$$

Given $E[U_{11} - E(U_{11})] = 0$, we just need to calculate $E\{[U_{11} - E(U_{11})]^2\}$:

$$E\{[U_{11} - E(U_{11})]^2\} = E\Bigg\{\frac{1}{Th^{1/2}} \sum_{s,r \in \mathbb{D}_1} \int K^2(\eta) d\eta \int_{\mathbb{R}^q} Re[\varepsilon_s(u)\varepsilon_r(u)^*] W(u) du$$

$$
\begin{aligned}
& -E\left[\frac{1}{Th^{1/2}}\sum_{s,r\in\mathbb{D}_1}\int K^2(\eta)d\eta\int_{\mathbb{R}^q} Re[\varepsilon_s(u)\varepsilon_r(u)^*]W(u)du\right]\Bigg\}^2 \\
= & \frac{1}{T^2h}\sum_{s,r\in\mathbb{D}_1}\sum_{l,m\in\mathbb{D}_1} E\left\{Re\left[\varpi(\Xi_s,\Xi_r)\varpi(\Xi_l,\Xi_m)\right]\right\}\left[\int K^2(\eta)d\eta\right]^2,
\end{aligned}
$$

where we define

$$
\varpi(\Xi_s,\Xi_r) = \int_{\mathbb{R}^q}[\varepsilon_s(u)\varepsilon_r(u)^* - E(\varepsilon_s(u)\varepsilon_r(u)^*)]W(u)du.
$$

By Lemma 1 of Yoshihara (1976) and Proposition 4 of Hong, Wang, and Wang (2014), $E\{[U_{11}-E(U_{11})]^2\} = op(T^{-1}h^{-1}p_T) = op(1)$. By Chebyshev's Inequality, we have $U_{11} = E(U_{11}) + op(1)$. Thus, we have $U_{11} = B_2 + op(1)$.

Next, we work on U_{12}. By Lipschitz condition, we have

$$
\left|K\left(\eta+\frac{s-r}{Th}\right) - K(\eta)\right| \le C\frac{|s-r|}{Th}.
$$

Thus,

$$
\begin{aligned}
E(U_{12}) &= \frac{2}{Th^{1/2}}\sum_{s,r\in\mathbb{D}_1}\int K(\eta)\left[K\left(\eta+\frac{s-r}{Th}\right)-K(\eta)\right]d\eta\int_{\mathbb{R}^q} Re\{E[\varepsilon_s(u)\varepsilon_r(u)^*]\}W(u)du \\
&\le C\frac{1}{Th^{3/2}}\sum_{j=1}^{p_T} j\beta(j)^{\frac{\delta}{1+\delta}} \\
&= o(1).
\end{aligned}
$$

By similar by tedious derivation as in showing $E\{U_{11}-[E(U_{11})]^2\} = o(1)$, we can show $E\{U_{12}-[E(U_{12})]^2\} = o(p_T^3T^{-3}h^{-3}) = o(1)$. Thus, $U_{12} = op(1)$ by Chebyshev's inequality.

Next, we work on U_2, which determines the asymptotic distribution of our test statistic:

$$
\begin{aligned}
U_2 &= \frac{1}{Th^{1/2}}\sum_{s,r\in\mathbb{D}_2}\int K(\eta)K\left(\eta+\frac{s-r}{Th}\right)d\eta\int_{\mathbb{R}^q} Re[\varepsilon_s(u)\varepsilon_t(u)^*]W(u)du \\
&= \frac{2}{Th^{1/2}}\sum_{s=1}^{T-p_T-1}\sum_{r=s+p_T+1}^{T}\int K(\eta)K\left(\eta+\frac{s-r}{Th}\right)d\eta\int_{\mathbb{R}^q} Re[\varepsilon_s(u)\varepsilon_r(u)^*]W(u)du
\end{aligned}
$$

.

Let's first look at the mean of U_2:

$$\begin{aligned} E(U_2) &= \frac{2}{Th^{1/2}} \sum_{j=p_T+1}^{T} (T-j-p_T) \int_{\mathbb{R}^q} E\left\{Re[\varepsilon_s(u)\varepsilon_{s+j}(u)^*]\right\} W(u)du \int K(\eta)K\left(\eta+\frac{j}{Th}\right)d\eta \\ &\leq Ch^{1/2} \sum_{j=p_T+1}^{T} j^2\beta(j) \iint K(\eta)K(\eta+\lambda)d\eta d\lambda \\ &\leq O(h^{1/2}p_T^{-1}) \\ &= o(1). \end{aligned}$$

Next, we work on the variance of U_2. Given $E(U_2) = o(1)$, we have $Var(U_2) = E(U_2^2) + op(1)$.

$$\begin{aligned} E(U_2^2) &= \frac{4}{T^2h} \sum_{s=1}^{T-p_T-1} \sum_{r=s+p_T+1}^{T} \sum_{l=1}^{T-p_T-1} \sum_{m=l+p_T+1}^{T} \int_{\mathbb{R}^{2q}} E\left\{Re[\varepsilon_s(u)\varepsilon_r(u)^*\varepsilon_l(v)\varepsilon_m(v)^*]\right\} W(u)W(v)dudv \\ &\quad \times \int K(\eta)K\left(\eta+\frac{r-s}{Th}\right)d\eta \int K(\lambda)K\left(\lambda+\frac{m-l}{Th}\right)d\lambda \\ &= \frac{4}{T^2h} \sum_{s=1}^{T-p_T-1} \sum_{r=s+p_T+1}^{T} \int_{\mathbb{R}^{2q}} E\left\{Re[\varepsilon_s(u)\varepsilon_r(u)^*\varepsilon_s(v)\varepsilon_r(v)^*]\right\} W(u)W(v)dudv \\ &\quad \times \left[\int K(\eta)K\left(\eta+\frac{r-s}{Th}\right)d\eta\right]^2 \\ &\quad + \frac{4}{T^2h} \sum_{s=1}^{T-p_T-1} \sum_{r=s+p_T+1}^{T} \sum_{m=s+p_T+1, r\neq m}^{T} \int_{\mathbb{R}^{2q}} E\left\{Re[\varepsilon_s(u)\varepsilon_r(u)^*\varepsilon_s(v)\varepsilon_m(v)^*]\right\} W(u)W(v)dudv \\ &\quad \times \int K(\eta)K\left(\eta+\frac{r-s}{Th}\right)d\eta \int K(\lambda)K\left(\lambda+\frac{m-s}{Th}\right)d\lambda \\ &\quad + \frac{4}{T^2h} \sum_{s\neq l}^{T-p_T-1} \sum_{r=max\{s,l\}+p_T+1}^{T} \int_{\mathbb{R}^{2q}} E\left\{Re[\varepsilon_s(u)\varepsilon_r(u)^*\varepsilon_l(v)\varepsilon_r(v)^*]\right\} W(u)W(v)dudv \\ &\quad \times \int K(\eta)K\left(\eta+\frac{r-s}{Th}\right)d\eta \int K(\lambda)K\left(\lambda+\frac{r-l}{Th}\right)d\lambda \\ &\quad + \frac{4}{T^2h} \sum_{s\neq l}^{T-p_T-1} \sum_{r=s+p_T+1}^{T} \sum_{m=l+p_T+1, m\neq r}^{T-l} \int_{\mathbb{R}^{2q}} E\left\{Re[\varepsilon_s(u)\varepsilon_r(u)^*\varepsilon_l(v)\varepsilon_m(v)^*]\right\} W(u)W(v)dudv \\ &\quad \times \int K(\eta)K\left(\eta+\frac{r-s}{Th}\right)d\eta \int K(\lambda)K\left(\lambda+\frac{m-l}{Th}\right)d\lambda \\ &= V_1 + V_2 + V_3 + V_4. \end{aligned}$$

Define

$$\sigma_{s,r}(u,v) = E\left[\varepsilon_s(u)\varepsilon_r(v)\right],$$

then we have

$$\begin{aligned}
& Re[\varepsilon_s(u)\varepsilon_r(u)^*\varepsilon_s(v)\varepsilon_r(v)^*] \\
= \;& Re\left\{[\sigma_{s,s}(u,v) + \varepsilon_{s,s}(u+v) - \varepsilon_s(u)\phi_s(v) - \varepsilon_s(v)\phi_s(u)]\right\} \\
& \times [\sigma_{r,r}(u,v) + \varepsilon_r(u+v) - \varepsilon_r(u)\phi_r(v) - \varepsilon_r(v)\phi_r(u)]^* \\
= \;& Re[\sigma_{s,s}(u,v)\sigma_{r,r}(u,v)^*] + Re\left\{\sigma_{s,s}(u,v)[\varepsilon_r(u+v) - \varepsilon_r(u)\phi_r(v) - \varepsilon_r(v)\phi_r(u)]^*\right\} \\
& + Re\left\{[\varepsilon_{s,s}(u+v) - \varepsilon_s(u)\phi_s(v) - \varepsilon_s(v)\phi_s(u)]\sigma_{r,r}(u,v)^*\right\} + Re[\Lambda_{s,r}(u,v)],
\end{aligned}$$

where

$$\begin{aligned}
\Lambda_{s,r}(u,v) = \;& \varepsilon_s(u+v)\varepsilon_r(u+v)^* + \phi_s(v)\phi_r(v)^*\varepsilon_s(u)\varepsilon_r(u)^* + \phi_s(u)\phi_r(u)^*\varepsilon_s(v)\varepsilon_r(v)^* \\
& + \phi_s(u)\phi_r(v)^*\varepsilon_s(v)\varepsilon_r(u)^* + \phi_s(v)\phi_r(u)^*\varepsilon_s(u)\varepsilon_r(v)^* - \phi_r(u)^*\varepsilon_{s,s}(u+v)\varepsilon_r(v)^* \\
& - \phi_r(v)^*\varepsilon_{s,s}(u+v)\varepsilon_r(u)^* - \phi_s(u)\varepsilon_s(v)\varepsilon_r(u+v)^* - \phi_s(v)\varepsilon_s(u)\varepsilon_r(u+v)^*,
\end{aligned}$$

which characterizes all the serial dependence.

$$\begin{aligned}
V_1 = \;& \frac{4}{T^2h}\sum_{s=1}^{T-p_T-1}\sum_{r=s+p_T+1}^{T}\int_{\mathbb{R}^{2q}} E\left\{Re[\varepsilon_s(u)\varepsilon_r(u)^*\varepsilon_s(v)\varepsilon_r(v)^*]\right\} W(u)W(v)dudv \\
& \times\left[\int K(\eta)K\left(\eta + \frac{r-s}{Th}\right)d\eta\right]^2 \\
= \;& \frac{4}{T^2h}\sum_{s=1}^{T-p_T-1}\sum_{j=p_T+1}^{T-s}\int_{\mathbb{R}^{2q}} E\left\{Re[\varepsilon_s(u)\varepsilon_{s+j}(u)^*\varepsilon_s(v)\varepsilon_{s+j}(v)^*]\right\} W(u)W(v)dudv \\
& \times\left[\int K(\eta)K\left(\eta + \frac{j}{Th}\right)d\eta\right]^2 \\
= \;& \frac{4}{T^2h}\sum_{s=1}^{T-p_T-1}\sum_{j=p_T+1}^{T-s}\int_{\mathbb{R}^{2q}} Re[\sigma_{s,s}(u,v)\sigma_{s+j,s+j}(u,v)^*]W(u)W(v)dudv\left[\int K(\eta)K\left(\eta + \frac{j}{Th}\right)d\eta\right]^2 \\
& + \frac{4}{T^2h}\sum_{s=1}^{T-p_T-1}\sum_{j=p_T+1}^{T-s}\int_{\mathbb{R}^{2q}} E\left\{Re[\Lambda_{s,s+j}(u,v)]\right\} W(u)W(v)dudv\left[\int K(\eta)K\left(\eta + \frac{j}{Th}\right)d\eta\right]^2
\end{aligned}$$

$$= \quad V_{11} + V_{12}, \text{ say.}$$

For V_{12},

$$\begin{aligned}
V_{12} &= \frac{4}{T^2 h} \sum_{s=1}^{T-p_T-1} \sum_{j=p_T+1}^{T-s} \int_{\mathbb{R}^{2q}} E\left\{Re[\Lambda_{s,s+j}(u,v)]\right\} W(u)W(v)dudv \left[\int K(\eta)K\left(\eta + \frac{j}{Th}\right)d\eta\right]^2 \\
&\leq \frac{4}{T^2 h} \sum_{j=p_T+1}^{T-1} (T-j)Cj^2\beta(j) \left[\int K(\eta)K\left(\eta + \frac{j}{Th}\right)d\eta\right]^2 \\
&\leq C \sum_{j=p_T}^{\infty} j^2\beta(j) \\
&= O(p_T^{-1}),
\end{aligned}$$

where the first inequality comes from the mixing condition, and the last inequality comes from the condition we imposed on p_T. For V_{11},

$$\begin{aligned}
V_{11} &= \frac{4}{T^2 h} \sum_{s=1}^{T-p_T-1} \sum_{j=p_T+1}^{T-s} \int_{\mathbb{R}^{2q}} Re[\sigma_{s,s}(u,v)\sigma_{s+j,s+j}(u,v)^*]W(u)W(v)dudv \left[\int K(\eta)K\left(\eta + \frac{j}{Th}\right)d\eta\right]^2 \\
&= \frac{4}{T^2 h} \sum_{j=p_T+1}^{T-1} (T-j) \int_{\mathbb{R}^{2q}} Re[\sigma_{1,1}(u,v)\sigma_{1+j,1+j}(u,v)^*]W(u)W(v)dudv \times \left[\int K(\eta)K\left(\eta + \frac{j}{Th}\right)d\eta\right]^2 \\
&= \frac{4}{Th} \sum_{j=p_T+1}^{T-1} \left(1 - \frac{j}{T}\right) \int_{\mathbb{R}^{2q}} Re[\sigma_{1,1}(u,v)\sigma_{1+j,1+j}(u,v)^*]W(u)W(v)dudv \times \left[\int K(\eta)K\left(\eta + \frac{j}{Th}\right)d\eta\right]^2.
\end{aligned}$$

If we have homoskedasticity, i.e., $Var(Y_t)$ is a constant, then we can further simplify V_{11} as:

$$V_{11} = 2\int_{\mathbb{R}^{2q}} |\sigma_{1,1}(u,v)|^2 W(u)W(v)dudv \int \left[\int K(\eta)K(\eta+\lambda)d\eta\right]^2 d\lambda + o(1),$$

where $\sigma_{1,1}(u,v) = E[\varepsilon_1(u)\varepsilon_1(v)] = \phi(u+v) - \phi(u)\phi(v)$.

Next, we work on V_2:

$$V_2 = \frac{4}{T^2 h} \sum_{s=1}^{T-p_T-1} \sum_{r=s+p_T+1}^{T} \sum_{m=s+p_T+1, r\neq m}^{T} \int_{\mathbb{R}^{2q}} E\left\{Re[\varepsilon_s(u)\varepsilon_r(u)^*\varepsilon_s(v)\varepsilon_m(v)^*]\right\} W(u)W(v)dudv$$

$$
\begin{aligned}
& \times \int K(\eta) K\left(\eta+\frac{r-s}{Th}\right) d\eta \int K(\lambda) K\left(\lambda+\frac{m-s}{Th}\right) d\lambda \\
= \; & \frac{4}{T^2 h} \sum_{s=1}^{T-p_T-1} \sum_{j=p_T+1}^{T-s} \sum_{k=p_T+1, j \neq k}^{T-s} \int_{\mathbb{R}^{2q}} E\left\{Re[\varepsilon_s(u)\varepsilon_{s+j}(u)^* \varepsilon_s(v)\varepsilon_{s+k}(v)^*]\right\} W(u)W(v)dudv \\
& \times \int K(\eta) K\left(\eta+\frac{j}{Th}\right) d\eta \int K(\lambda) K\left(\lambda+\frac{k}{Th}\right) d\lambda \\
= \; & \frac{4}{T^2 h} \sum_{s=1}^{T-p_T-1} \sum_{j=p_T+1}^{T-s} \sum_{k=p_T+1, j \neq k}^{T-s} \int_{\mathbb{R}^{2q}} Re\left\{\sigma_{s,s}(u,v) E[\varepsilon_{s+j}(u)\varepsilon_{s+k}(v)]^*\right\} W(u)W(v)dudv \\
& \times \int K(\eta) K\left(\eta+\frac{j}{Th}\right) d\eta \int K(\lambda) K\left(\lambda+\frac{k}{Th}\right) d\lambda \\
& + \frac{4}{T^2 h} \sum_{s=1}^{T-p_T-1} \sum_{j=p_T+1}^{T-s} \sum_{k=p_T+1, j \neq k}^{T-s} \int_{\mathbb{R}^{2q}} Re\left\{E[\varepsilon_{s,s}(u+v)\varepsilon_{s+j}(u)^* \varepsilon_{s+k}(v)^*]\right\} W(u)W(v)dudv \\
& \times \int K(\eta) K\left(\eta+\frac{j}{Th}\right) d\eta \int K(\lambda) K\left(\lambda+\frac{k}{Th}\right) d\lambda \\
& + \frac{4}{T^2 h} \sum_{s=1}^{T-p_T-1} \sum_{j=p_T+1}^{T-s} \sum_{k=p_T+1, j \neq k}^{T-s} \int_{\mathbb{R}^{2q}} Re\left\{E[\varepsilon_s(u)\phi_s(v)\varepsilon_{s+j}(u)^* \varepsilon_{s+k}(v)^*]\right\} W(u)W(v)dudv \\
& \times \int K(\eta) K\left(\eta+\frac{j}{Th}\right) d\eta \int K(\lambda) K\left(\lambda+\frac{k}{Th}\right) d\lambda \\
& + \frac{4}{T^2 h} \sum_{s=1}^{T-p_T-1} \sum_{j=p_T+1}^{T-s} \sum_{k=p_T+1, j \neq k}^{T-s} \int_{\mathbb{R}^{2q}} Re\left\{E[\phi_s(u)\varepsilon_s(v)\varepsilon_{s+j}(u)^* \varepsilon_{s+k}(v)^*]\right\} W(u)W(v)dudv \\
& \times \int K(\eta) K\left(\eta+\frac{j}{Th}\right) d\eta \int K(\lambda) K\left(\lambda+\frac{k}{Th}\right) d\lambda \\
= \; & V_{21} + V_{22} + V_{23} + V_{24}.
\end{aligned}
$$

For V_{21}, we can split it into two parts:

$$
\begin{aligned}
V_{21} \; = \; & \frac{4}{T^2 h} \sum_{s=1}^{T-p_T-1} \sum_{j=p_T+1}^{T-s} \sum_{k=p_T+1, j \neq k}^{T-s} \int_{\mathbb{R}^{2q}} Re\left\{\sigma_{s,s}(u,v) E[\varepsilon_{s+j}(u)\varepsilon_{s+k}(v)]^*\right\} W(u)W(v)dudv \\
& \times \int K(\eta) K\left(\eta+\frac{j}{Th}\right) d\eta \int K(\lambda) K\left(\lambda+\frac{k}{Th}\right) d\lambda \\
= \; & \frac{4}{T^2 h} \sum_{s=1}^{T-p_T-1} \sum_{j,k=p_T+1, 0<|j-k|\leq p_T}^{T-s} \int_{\mathbb{R}^{2q}} Re\left\{\sigma_{s,s}(u,v) E[\varepsilon_{s+j}(u)\varepsilon_{s+k}(v)]^*\right\} W(u)W(v)dudv \\
& \times \int K(\eta) K\left(\eta+\frac{j}{Th}\right) d\eta \int K(\lambda) K\left(\lambda+\frac{k}{Th}\right) d\lambda
\end{aligned}
$$

$$
\begin{aligned}
& +\frac{4}{T^2 h}\sum_{s=1}^{T-p_T-1}\sum_{j,k=p_T+1,|j-k|>p_T}^{T-s}\int_{\mathbb{R}^{2q}} Re\left\{\sigma_{s,s}(u,v)E[\varepsilon_{s+j}(u)\varepsilon_{s+k}(v)]^*\right\}W(u)W(v)dudv \\
& \times\int K(\eta)K\left(\eta+\frac{j}{Th}\right)d\eta\int K(\lambda)K\left(\lambda+\frac{k}{Th}\right)d\lambda \\
= \;& V_{211}+V_{212}.
\end{aligned}
$$

We claim that $V_{212}=op(1)$, since $\sum_{|j|>p_T}E[\varepsilon_s(u)^*\varepsilon_{s+j}(v)^*]<p_T^{-1}$ given the conditions on p_T. Thus, we just need to calculate V_{211}:

$$
\begin{aligned}
V_{211} &= \frac{4}{T^2 h}\sum_{s=1}^{T-p_T-1}\sum_{j,k=p_T+1,0<|j-k|\leq p_T}^{T-s}\int_{\mathbb{R}^{2q}} Re\{\sigma_{s,s}(u,v)E[\varepsilon_{s+j}(u)^*\varepsilon_{s+k}(v)^*]\}W(u)W(v)dudv \\
&\quad \times\int K(\eta)K\left(\eta+\frac{j}{Th}\right)d\eta\int K(\lambda)K\left(\lambda+\frac{k}{Th}\right)d\lambda \\
&= \frac{4}{T^2 h}\sum_{s=1}^{T-p_T-1}\sum_{j=p_T+1}^{T-s}\int_{\mathbb{R}^{2q}} Re\Bigg\{\sigma_{s,s}(u,v) \\
&\quad \times\sum_{l=1}^{\min\{T-s-j,p_T\}} E[\varepsilon_{s+j}(u)\varepsilon_{s+j+l}(v)+\varepsilon_{s+j}(v)\varepsilon_{s+j+l}(u)]^*\Bigg\}W(u)W(v)dudv \\
&\quad \times\int K(\eta)K\left(\eta+\frac{j}{Th}\right)d\eta\int K(\lambda)K\left(\lambda+\frac{j+l}{Th}\right)d\lambda \\
&= \frac{4}{T^2 h}\sum_{s=1}^{T-p_T-1}\sum_{j=p_T+1}^{T-s}\int_{\mathbb{R}^{2q}} Re\Bigg\{\sigma_{s,s}(u,v) \\
&\quad \times\sum_{l=1}^{\min\{T-s-j,p_T\}} [\sigma_{s+j,s+j+l}(u,v)+\sigma_{s+j+l,s+j}(u,v)]^*\Bigg\}W(u)W(v)dudv \\
&\quad \times\left[\int K(\eta)K\left(\eta+\frac{j}{Th}\right)d\eta\right]^2+o(1)
\end{aligned}
$$

Under homoskedasticity, we can simplify V_{211} as

$$
\begin{aligned}
V_{211} &= 2\int_{\mathbb{R}^{2q}} Re\left\{\sigma_{1,1}(u,v)\sum_{l=1}^{p_T}\left[\sigma_{1,1+l}(u,v)^*+\sigma_{1+l,1}(u,v)^*\right]\right\}W(u)W(v)dudv \\
&\quad \times\int\left[\int K(\eta)K(\eta+\lambda)d\eta\right]^2 d\lambda+o(1)
\end{aligned}
$$

For V_{22}, V_{23}, and V_{24}, we can show they are higher order terms using the following condition derived from Roussas and Ioannides (1987):

$sup_{u,v}E[\varepsilon_s(u)\varepsilon_r(u)\varepsilon_l(u)^*] \leq C\beta(d)^{\delta/(1+\delta)}$. The proof is quite tedious and we skip it for space.

Next, we work on V_3. Following the analogous analysis of V_2, we can show that

$$\begin{aligned}
V_3 &= \frac{4}{T^2h}\sum_{s\neq l}^{T-p_T-1}\sum_{r=max\{s,l\}+p_T+1}^{T}\int_{\mathbb{R}^{2q}} E\left\{Re[\varepsilon_s(u)\varepsilon_r(u)^*\varepsilon_l(v)\varepsilon_r(v)^*]\right\}W(u)W(v)dudv \\
&\quad \times\int K(\eta)K\left(\eta+\frac{r-s}{Th}\right)d\eta\int K(\lambda)K\left(\lambda+\frac{r-l}{Th}\right)d\lambda \\
&= \frac{4}{T^2h}\sum_{s=1}^{T-p_T-1}\sum_{j=p_T+1}^{T-s}\int_{\mathbb{R}^{2q}} Re\Bigg\{\sum_{k=1}^{min\{p_T,T-p_T-1-s\}}[\sigma_{s,s+k}(u,v)+\sigma_{s+k,s}(u,v)] \\
&\quad \times\sigma_{s+j,s+j}(u,v)^*\Bigg\}W(u)W(v)dudv\left[\int K(\eta)K\left(\eta+\frac{j}{Th}\right)d\eta\right]^2+o(1).
\end{aligned}$$

At last, we work on V_4:

$$\begin{aligned}
V_4 &= \frac{4}{T^2h}\sum_{s\neq l}^{T-p_T-1}\sum_{r=s+p_T+1}^{T}\sum_{m=l+p_T+1,r\neq m}^{T}\int_{\mathbb{R}^{2q}} E\left\{Re[\varepsilon_s(u)\varepsilon_r(u)^*\varepsilon_l(v)\varepsilon_m(v)^*]\right\}W(u)W(v)dudv \\
&\quad \times\int K(\eta)K\left(\eta+\frac{r-s}{Th}\right)d\eta\int K(\lambda)K\left(\lambda+\frac{m-l}{Th}\right)d\lambda \\
&= \frac{4}{T^2h}\sum_{s\neq l}^{T-p_T-1}\sum_{r=s+p_T+1}^{T}\sum_{m=l+p_T+1,r\neq m}^{T}\int_{\mathbb{R}^{2q}} Re\left\{E[\varepsilon_s(u)\varepsilon_l(v)]E[\varepsilon_r(u)^*\varepsilon_m(v)^*]\right\}W(u)W(v)dudv \\
&\quad \times\int K(\eta)K\left(\eta+\frac{r-s}{Th}\right)d\eta\int K(\lambda)K\left(\lambda+\frac{m-l}{Th}\right)d\lambda \\
&\quad +\frac{4}{T^2h}\sum_{s\neq l}^{T-p_T-1}\sum_{r=s+p_T+1}^{T}\sum_{m=l+p_T+1,r\neq m}^{T}\int_{\mathbb{R}^{2q}} Re\{E[\varepsilon_s(u)\varepsilon_r(u)^*\varepsilon_l(v)\varepsilon_m(v)^*]-E[\varepsilon_s(u)\varepsilon_l(v)] \\
&\quad \times E[\varepsilon_r(u)^*\varepsilon_m(v)^*]\}W(u)W(v)dudv\int K(\eta)K\left(\eta+\frac{r-s}{Th}\right)d\eta\int K(\lambda)K\left(\lambda+\frac{m-l}{Th}\right)d\lambda \\
&= V_{41}+V_{42}.
\end{aligned}$$

For V_{41}, we have

$$V_{41} = \frac{4}{T^2h}\int_{\mathbb{R}^{2q}} Re\left\{\sum_{s=1}^{T-p_T-1}\sum_{l=1,l\neq s}^{T-p_T-1}E[\varepsilon_s(u)\varepsilon_l(v)]\sum_{r=s+p_T+1}^{T}\sum_{\substack{m=l+p_T+1\\ m\neq r}}^{T}E[\varepsilon_r(u)^*\varepsilon_m(v)^*]\right\}W(u)W(v)dudv$$

$$
\times \int K(\eta)K\left(\eta + \frac{r-s}{Th}\right)d\eta \int K(\lambda)K\left(\lambda + \frac{m-l}{Th}\right)d\lambda
$$

$$
\begin{aligned}
= \;& \frac{4}{T^2h}\int_{\mathbb{R}^{2q}} Re\left\{\sum_{s,l:|s-l|>p_T}^{T-p_T-1} E[\varepsilon_s(u)\varepsilon_l(v)] \sum_{r=s+p_T+1}^{T} \sum_{\substack{m=l+p_T+1\\|m-r|>p_T}}^{T} E[\varepsilon_r(u)^*\varepsilon_m(v)^*]\right\} W(u)W(v)dudv \\
& \times \int K(\eta)K\left(\eta + \frac{r-s}{Th}\right)d\eta \int K(\lambda)K\left(\lambda + \frac{m-l}{Th}\right)d\lambda \\
& + \frac{4}{T^2h}\int_{\mathbb{R}^{2q}} Re\left\{\sum_{s,l:|s-l|>p_T}^{T-p_T-1} E[\varepsilon_s(u)\varepsilon_l(v)] \sum_{r=s+p_T+1}^{T} \sum_{\substack{m=l+p_T+1\\|m-r|\le p_T}}^{T} E[\varepsilon_r(u)^*\varepsilon_m(v)^*]\right\} W(u)W(v)dudv \\
& \times \int K(\eta)K\left(\eta + \frac{r-s}{Th}\right)d\eta \int K(\lambda)K\left(\lambda + \frac{m-l}{Th}\right)d\lambda \\
& + \frac{4}{T^2h}\int_{\mathbb{R}^{2q}} Re\left\{\sum_{s,l:|s-l|\le p_T}^{T-p_T-1} E[\varepsilon_s(u)\varepsilon_l(v)] \sum_{r=s+p_T+1}^{T} \sum_{\substack{m=l+p_T+1\\|m-r|>p_T}}^{T} E[\varepsilon_r(u)^*\varepsilon_m(v)^*]\right\} W(u)W(v)dudv \\
& \times \int K(\eta)K\left(\eta + \frac{r-s}{Th}\right)d\eta \int K(\lambda)K\left(\lambda + \frac{m-l}{Th}\right)d\lambda \\
& + \frac{4}{T^2h}\int_{\mathbb{R}^{2q}} Re\left\{\sum_{s,l:|s-l|\le p_T}^{T-p_T-1} E[\varepsilon_s(u)\varepsilon_l(v)] \sum_{r=s+p_T+1}^{T} \sum_{\substack{m=l+p_T+1\\|m-r|\le p_T}}^{T} E[\varepsilon_r(u)^*\varepsilon_m(v)^*]\right\} W(u)W(v)dudv \\
& \times \int K(\eta)K\left(\eta + \frac{r-s}{Th}\right)d\eta \int K(\lambda)K\left(\lambda + \frac{m-l}{Th}\right)d\lambda \\
= \;& V_{411} + V_{412} + V_{413} + V_{414}.
\end{aligned}
$$

For V_{411}, V_{412}, and V_{413}, we can show that they are all higher order terms that converge to 0 faster than V_{414} given $\sum_{j>p_T} E[\varepsilon_s(u)\varepsilon_{s+j}(u)] < p_T^{-1}$. So we just need to calculate V_{414}:

$$
\begin{aligned}
V_{414} = \;& \frac{4}{T^2h}\int_{\mathbb{R}^{2q}} Re\left\{\sum_{s,l:|s-l|\le p_T}^{T-p_T-1} E[\varepsilon_s(u)\varepsilon_l(v)] \sum_{r=s+p_T+1}^{T} \sum_{\substack{m=l+p_T+1\\|m-r|\le p_T}}^{T} E[\varepsilon_r(u)^*\varepsilon_m(v)^*]\right\} W(u)W(v)dudv \\
& \times \int K(\eta)K\left(\eta + \frac{r-s}{Th}\right)d\eta \int K(\lambda)K\left(\lambda + \frac{m-l}{Th}\right)d\lambda \\
= \;& \frac{4}{T^2h}\int_{\mathbb{R}^{2q}} Re\Bigg\{\sum_{s=1}^{T-p_T-1} \sum_{k=1}^{min\{T-p_T-1-s,p_T\}} E[\varepsilon_s(u)\varepsilon_{s+k}(v) + \varepsilon_{s+k}(u)\varepsilon_s(v)]
\end{aligned}
$$

$$
\begin{aligned}
&\times \sum_{r=s+p_T+1}^{T} \sum_{k'=1}^{min\{T-r,p_T\}} E[\varepsilon_r(u)^* \varepsilon_{r+k'}(v)^* + \varepsilon_{r+k'}(u)^* \varepsilon_r(v)^*] \Big\} W(u)W(v)dudv \\
&\times \int K(\eta)K\left(\eta + \frac{r-s}{Th}\right)d\eta \int K(\lambda)K\left(\lambda + \frac{r+k'-(s+k)}{Th}\right)d\lambda \\
= \;& \frac{4}{T^2h} \sum_{s=1}^{T-p_T-1} \sum_{j=p_T+1}^{T-s} \int_{\mathbb{R}^{2q}} Re\Big\{ \sum_{k=1}^{min\{T-p_T-1-s,p_T\}} E[\varepsilon_s(u)\varepsilon_{s+k}(v) + \varepsilon_{s+k}(u)\varepsilon_s(v)] \\
&\times \sum_{k'=1}^{min\{T-s-j,p_T\}} E[\varepsilon_{s+j}(u)^* \varepsilon_{s+j+k'}(v)^* + \varepsilon_{s+j+k'}(u)^* \varepsilon_{s+j}(v)^*] \Big\} W(u)W(v)dudv \\
&\times \int K(\eta)K\left(\eta + \frac{j}{Th}\right)d\eta \int K(\lambda)K\left(\lambda + \frac{j}{Th} + \frac{k'-k}{Th}\right)d\lambda \\
= \;& \frac{4}{T^2h} \sum_{s=1}^{T-p_T-1} \sum_{j=p_T+1}^{T-s} \int_{\mathbb{R}^{2q}} Re\Big\{ \sum_{k=1}^{min\{T-p_T-1-s,p_T\}} [\sigma_{s,s+k}(u,v) + \sigma_{s+k,s}(u,v)] \\
&\times \sum_{k'=1}^{min\{T-s-j,p_T\}} [\sigma_{s+j,s+j+k'}(u,v)^* + \sigma_{s+j+k',s+j}(u,v)^*] \Big\} W(u)W(v)dudv \\
&\times \left[\int K(\eta)K\left(\eta + \frac{j}{Th}\right)d\eta\right]^2 + o(1).
\end{aligned}
$$

Under homoskedasticity, we can further simplify V_{414} as

$$
V_{414} = 2\int_{\mathbb{R}^{2q}} \left| \sum_{k=1}^{p_T} [\sigma_{1,1+k}(u,v) + \sigma_{1+k,1}(u,v)] \right|^2 W(u)W(v)dudv \int \left[\int K(\eta)K(\eta+\lambda)d\eta\right]^2 d\lambda + o(1).
$$

At last, we have to show $V_{42} = op(1)$:

$$
\begin{aligned}
V_{42} = \;& \frac{4}{T^2h} \sum_{s \neq l}^{T-p_T-1} \sum_{r=s+p_T+1}^{T} \sum_{m=l+p_T+1, m\neq r}^{T} \int_{\mathbb{R}^{2q}} Re\{E[\varepsilon_s(u)\varepsilon_r(u)^* \varepsilon_l(v)\varepsilon_m(v)^*] - E[\varepsilon_s(u)\varepsilon_l(v)] \\
&\times E[\varepsilon_r(u)^* \varepsilon_m(v)^*]\} W(u)W(v)dudv \int K(\eta)K\left(\eta + \frac{r-s}{Th}\right)d\eta \int K(\lambda)K\left(\lambda + \frac{m-l}{Th}\right)d\lambda \\
= \;& \frac{4}{T^2h} \sum_{s \neq l}^{T-p_T-1} \sum_{r=s+p_T+1}^{T} \sum_{m=l+p_T+1, m\neq r}^{T} \int_{\mathbb{R}^{2q}} Re\{Cov[\varepsilon_s(u)\varepsilon_l(v), \varepsilon_r(u)^* \varepsilon_m(v)^*]\} W(u)W(v)dudv \\
& \int K(\eta)K\left(\eta + \frac{r-s}{Th}\right)d\eta \int K(\lambda)K\left(\lambda + \frac{m-l}{Th}\right)d\lambda \\
\leq \;& C\frac{p_T}{Th} \sum_{j=1}^{T} \beta(j)^{\frac{\delta}{1+\delta}} \\
= \;& o(1),
\end{aligned}
$$

where the inequality comes from the mixing condition and the definition of

p_T which is analogous to the proof of $Var(U_{11}) = o(1)$. After we obtain the expressions for V_1, V_2, V_3, and V_4, we have

$$\begin{aligned}
V &= V_1 + V_2 + V_3 + V_4 \\
&= \frac{4}{T^2 h} \sum_{s=1}^{T-p_T-1} \sum_{j=p_T+1}^{T-s} \int_{\mathbb{R}^{2q}} Re[\sigma_{s,s}(u,v)\sigma_{s+j,s+j}(u,v)^*] W(u)W(v)dudv \left[\int K(\eta)K\left(\eta + \frac{j}{Th}\right)d\eta\right]^2 \\
&\quad + \frac{4}{T^2 h} \sum_{s=1}^{T-p_T-1} \sum_{j=p_T+1}^{T-s} \int_{\mathbb{R}^{2q}} Re\Bigg\{\sigma_{s,s}(u,v) \\
&\quad \times \sum_{l=1}^{\min\{T-s-j,p_T\}} [\sigma_{s+j,s+j+l}(u,v) + \sigma_{s+j+l,s+j}(u,v)]^* \Bigg\} W(u)W(v)dudv \\
&\quad \times \left[\int K(\eta)K\left(\eta + \frac{j}{Th}\right)d\eta\right]^2 \\
&\quad + \frac{4}{T^2 h} \sum_{s=1}^{T-p_T-1} \sum_{j=p_T+1}^{T-s} \int_{\mathbb{R}^{2q}} Re\Bigg\{ \sum_{k=1}^{min\{p_T, T-p_T-1-s\}} [\sigma_{s,s+k}(u,v) + \sigma_{s+k,s}(u,v)] \\
&\quad \times \sigma_{s+j,s+j}(u,v)^* \Bigg\} W(u)W(v)dudv \left[\int K(\eta)K\left(\eta + \frac{j}{Th}\right)d\eta\right]^2 \\
&\quad + \frac{4}{T^2 h} \sum_{s=1}^{T-p_T-1} \sum_{j=p_T+1}^{T-s} \int_{\mathbb{R}^{2q}} Re\Bigg\{ \sum_{k=1}^{min\{T-p_T-1-s, p_T\}} [\sigma_{s,s+k}(u,v) + \sigma_{s+k,s}(u,v)] \\
&\quad \times \sum_{k'=1}^{min\{T-s-j,p_T\}} [\sigma_{s+j,s+j+k'}(u,v)^* + \sigma_{s+j+k',s+j}(u,v)^*] \Bigg\} W(u)W(v)dudv \\
&\quad \times \left[\int K(\eta)K\left(\eta + \frac{j}{Th}\right)d\eta\right]^2 + o(1). \\
&= \frac{4}{T^2 h} \sum_{s=1}^{T-p_T-1} \sum_{j=p_T+1}^{T-s} \left[\int K(\eta)K\left(\eta + \frac{j}{Th}\right)d\eta\right]^2 \\
&\quad \int_{\mathbb{R}^{2q}} Re\left\{\left[\sum_{k=-min\{T-p_T-1-s,p_T\}}^{min\{T-p_T-1-s,p_T\}} \sigma_{s,s+k}(u,v)\right]\left[\sum_{k'=-min\{T-s-j,p_T\}}^{min\{T-s-j,p_T\}} \sigma_{s+j,s+j+k'}(u,v)\right]^*\right\} \\
&\quad \times W(u)W(v)dudv + op(1) \\
&= 2\int_{\mathbb{R}^{2q}} |\Omega_1(u,v)|^2 W(u)W(v)dudv \int \left[\int K(\eta)K(\eta+\lambda)d\eta\right]^2 d\lambda + op(1).
\end{aligned}$$

Finally, we have to prove asymptotic normality of $Z \equiv U_2/\sqrt{Var(U_2)} \xrightarrow{d} N(0,1)$. Here, we make use of the central limit theorem used by Hong, Wang, and Wang (2014) and we can show that for all $a \in \mathbb{R}$, the moment generating function of Z is

$$M_Z(a) = E(e^{aZ}) = \sum_{r=1}^{\infty} \frac{1}{r!}\left(\frac{a^2}{2}\right)^r = e^{\frac{a^2}{2}}.$$

By the uniqueness of moment generating function, we can show that $Z \sim N(0,1)$.

□

Proof of Proposition A.1.3. We want to show that the estimators $\hat{B}$ and $\hat{V}$ are consistent estimators for B and V respectively. The proof of $\hat{B} - B = op(1)$ and $\hat{V} - V = op(1)$ are similar. Given

$$\hat{B} = h^{-1/2} \int_{\mathbb{R}^q} |\hat{\Omega}_1(u,u)| W(u) du \int K^2(\eta) d\eta,$$

and

$$B = h^{-1/2} \int_{\mathbb{R}^q} |\Omega_1(u,u)| W(u) du \int K^2(\eta) d\eta,$$

it is sufficient to show that $\hat{\Omega}_1(u,u) - \Omega_1(u,u) = op(h^{1/2})$. Let $\bar{\Omega}_1(u,u) = \sum_{j=-p_T}^{p_T} k(\frac{j}{p_T})\bar{\sigma}_{1,1+j}(u,u)$ where $\bar{\sigma}_{1,1+j}(u,u) = \frac{1}{T-|j|}\sum_{t=1}^{T-|j|} \varepsilon_t(u)\varepsilon_{t+j}(u)$. Then we can decompose

$$\begin{aligned}
& \hat{\Omega}_1(u,u) - \Omega_1(u,u) \\
= \;& \bar{\Omega}_1(u,u) - \Omega_1(u,u) + \hat{\Omega}_1(u,u) - \bar{\Omega}_1(u,u) \\
= \;& D_1 + D_2.
\end{aligned}$$

First, we can decompose $D_1 = E(D_1) + D_1 - E(D_1)$.

$$\begin{aligned}
E(D_1) &= E[\bar{\Omega}_1(u,u) - \Omega_1(u,u)] \\
&= E\left[\sum_{j=-p_T}^{p_T} k\left(\frac{j}{p_T}\right)\bar{\sigma}_{1,1+j}(u,u) - \sum_{j=-\infty}^{\infty}\sigma_{1,1+j}(u,u)\right] \\
&= \sum_{j=-p_T}^{p_T} k\left(\frac{j}{p_T}\right)\sigma_{1,1+j}(u,u) - \sum_{j=-\infty}^{\infty}\sigma_{1,1+j}(u,u)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{j=-p_T}^{p_T} \left[k\left(\frac{j}{p_T}\right) - 1 \right] \sigma_{1,1+j}(u,u) - \sum_{|j|>p_T} \sigma_{1,1+j}(u,u) \\
&= ED_{11} + ED_{12}
\end{aligned}
$$

We show that $ED_{11} = O(p_T^{-1})$ by Lipschitz condition. And $ED_{12} = O(p_T^{-1})$ by the condition on p_T: $\sum_{|j|>p_T} \sigma_{1,1+j}(u,u) \leq p_T^{-1}$. Therefore, we have $ED_1 = op(h^{1/2})$ since $h^{1/2} p_T \to \infty$. For $D_1 - ED_1$, we can follow Hong et al. (2014) to show that $ED_1^2 = O(T^{-1}) = o(h^{1/2})$. Thus, $D_1 = op(h^{1/2})$ by Chebyshev's inequality.

Now, we will work on D_2:

$$
\begin{aligned}
D_2 &= \hat{\Omega}_1(u,u) - \bar{\Omega}_1(u,u) \\
&= \sum_{j=-p_T}^{p_T} k\left(\frac{j}{p_T}\right) \frac{1}{T-|j|} \sum_{t=1}^{T-|j|} [\hat{\varepsilon}_t(u)\hat{\varepsilon}_{t+j}(u) - \varepsilon_t(u)\varepsilon_{t+j}(u)] \\
&= \sum_{j=-p_T}^{p_T} k\left(\frac{j}{p_T}\right) \frac{1}{T-|j|} \sum_{t=1}^{T-|j|} [\hat{\varepsilon}_t(u)\hat{\varepsilon}_{t+j}(u) - \hat{\varepsilon}_t(u)\varepsilon_{t+j}(u) + \hat{\varepsilon}_t(u)\varepsilon_{t+j}(u) - \varepsilon_t(u)\varepsilon_{t+j}(u)] \\
&= \sum_{j=-p_T}^{p_T} k\left(\frac{j}{p_T}\right) \frac{1}{T-|j|} \sum_{t=1}^{T-|j|} [\hat{\varepsilon}_t(u)\{\hat{\varepsilon}_{t+j}(u) - \varepsilon_{t+j}(u)\}] \\
&\quad + \sum_{j=-p_T}^{p_T} k\left(\frac{j}{p_T}\right) \frac{1}{T-|j|} \sum_{t=1}^{T-|j|} [\{\hat{\varepsilon}_t(u) - \varepsilon_t(u)\}\varepsilon_{t+j}(u)] \\
&= \sum_{j=-p_T}^{p_T} k\left(\frac{j}{p_T}\right) \frac{1}{T-|j|} \sum_{t=1}^{T-|j|} [\hat{\varepsilon}_t(u) - \varepsilon_t(u)][\hat{\varepsilon}_{t+j}(u) - \varepsilon_{t+j}(u)] \\
&\quad + \sum_{j=-p_T}^{p_T} k\left(\frac{j}{p_T}\right) \frac{1}{T-|j|} \sum_{t=1}^{T-|j|} [\varepsilon_t(u)\{\hat{\varepsilon}_{t+j}(u) - \varepsilon_{t+j}(u)\}] \\
&\quad + \sum_{j=-p_T}^{p_T} k\left(\frac{j}{p_T}\right) \frac{1}{T-|j|} \sum_{t=1}^{T-|j|} [\{\hat{\varepsilon}_t(u) - \varepsilon_t(u)\}\varepsilon_{t+j}(u)] \\
&= D_{21} + D_{22} + D_{23}.
\end{aligned}
$$

Under $\mathbb{H}_0$, $\hat{\varepsilon}_t(u) - \varepsilon_t(u) = \phi(u) - \hat{\phi}(u) = Op(T^{-1/2})$ for all t. Thus, we have $D_{21} = Op(T^{-1}p_T), D_{22} = Op(T^{-1/2}p_T)$, and $D_{22} = Op(T^{-1/2}p_T)$. Then we know that $D_2 = op(h^{1/2})$. Thus, we have shown that $\hat{\Omega}_1(u,u) - \Omega_1(u,u) = op(h^{1/2})$ and therefore $\hat{B} - B = op(1)$. □

Proof of Theorem A.1.2. Under $\mathbb{H}_0 : \phi_t(u) = \phi_0(u)$, we can decompose $\hat{Q}_2$ as

$$
\begin{aligned}
\hat{Q}_2 &= h^{1/2} \sum_{t=1}^{T} \int_{\mathbb{R}^q} \left| \hat{\phi}_0(u) - \phi_0(u) \right|^2 W(u)du \\
&= h^{1/2} \sum_{t=1}^{T} \int_{\mathbb{R}^q} \left| \frac{1}{T} \sum_{s=1}^{T} (Y_s e^{iu'Z_s} - \phi_0(u)) \right|^2 W(u)du \\
&= h^{1/2} \sum_{t=1}^{T} \int_{\mathbb{R}^q} \left| \frac{1}{T} \sum_{s=1}^{T} \varepsilon_s(u) \right|^2 W(u)du \\
&= T^{-1} h^{1/2} \sum_{s=1}^{T} \int_{\mathbb{R}^q} |\varepsilon_s(u)|^2 W(u)du + T^{-1} h^{1/2} \sum_{s \neq r}^{T} \int_{\mathbb{R}^q} Re\left[\varepsilon_s(u)\varepsilon_r(u)^*\right] W(u)du \\
&= \hat{Q}_{21} + \hat{Q}_{22}.
\end{aligned}
$$

It is straightforward to show that $E\hat{Q}_{21} = Op(h^{1/2}) = op(1)$. For $var(\hat{Q}_{21})$:

$$
\begin{aligned}
var(\hat{Q}_{21}) &= T^{-2} h \sum_{s=1}^{T} var\left(\int_{\mathbb{R}^q} |\varepsilon_s(u)|^2 W(u)du \right) \\
&\quad + T^{-2} h \sum_{s \neq r}^{T} cov\left(\int_{\mathbb{R}^q} |\varepsilon_s(u)|^2 W(u)du, \int_{\mathbb{R}^q} |\varepsilon_r(u)|^2 W(u)du \right) \\
&= T^{-1} h var\left(\int_{\mathbb{R}^q} |\varepsilon_1(u)|^2 W(u)du \right) \\
&\quad + 2T^{-1} h \sum_{j=1}^{T-1} \left(1 - \frac{j}{T}\right) cov\left(\int_{\mathbb{R}^q} |\varepsilon_1(u)|^2 W(u)du, \int_{\mathbb{R}^q} \left|\varepsilon_{1+j}(u)\right|^2 W(u)du \right) \\
&= O(T^{-1}h) = o(1),
\end{aligned}
$$

by mixing conditions. By Chebyshev's inequality, we have $\hat{Q}_{21} = op(1)$.

For $\hat{Q}_{22}$:

$$
\begin{aligned}
E(\hat{Q}_{22}) &= 2T^{-1} h^{1/2} \sum_{1 \leq s < r \leq T} \int_{\mathbb{R}^q} Re[cov(\varepsilon_s(u), \varepsilon_r(u))] W(u)du \\
&= 2h^{1/2} \sum_{j=1}^{T-1} \left(1 - \frac{j}{T}\right) \int_{\mathbb{R}^q} Re[cov(\varepsilon_1(u), \varepsilon_{1+j}(u))] W(u)du \\
&\leq 2h^{1/2} C \sum_{j=1}^{\infty} \beta(j)^{\frac{\delta}{1+\delta}} \\
&= O(h^{1/2}).
\end{aligned}
$$

Furthermore, by Lemma A(ii) in Hjellvik et al. (1998),

$$\begin{aligned}
E(\hat{Q}_{22}^2) &= T^{-2}hE\left[\sum_{s\neq r}^{T}\int_{\mathbb{R}^q} Re\left[\varepsilon_s(u)\varepsilon_r(u)^*\right]W(u)du\right]^2 \\
&= T^{-2}hE\left\{\sum_{1\leq s<r\leq T}\left[\int_{\mathbb{R}^q} Re\left[\varepsilon_s(u)\varepsilon_r(u)^*\right]W(u)du + \int_{\mathbb{R}^q} Re\left[\varepsilon_r(u)\varepsilon_s(u)^*\right]W(u)du\right]\right\}^2 \\
&\leq hC\max_{1\leq j\leq T-1} E\left\{\left[\int_{\mathbb{R}^q} Re\left[\varepsilon_1(u)\varepsilon_{1+j}(u)^*\right]W(u)du + \int_{\mathbb{R}^q} Re\left[\varepsilon_{1+j}(u)\varepsilon_1(u)^*\right]W(u)du\right]^{2(1+\delta)}\right\}^{\frac{1}{1+\delta}} \\
&\quad \times \sum_{j=1}^{\infty} j\beta(j)^{\frac{\delta}{1+\delta}} \\
&= O(h).
\end{aligned}$$

By Chebyshev's Inequality, $\hat{Q}_{22} = op(1)$. Therefore, we have shown $\hat{Q}_2 = op(1)$. Intuitively, $\hat{Q}_2$ represents the error introduced by parametric estimation which should converge to 0 faster than nonparametric estimation. Thus, $\hat{Q}_2$ converges to 0 faster than $\hat{Q}_1$. □

Proof of Theorem A.1.3. Under the $\mathbb{H}_0$, we can decompose $\hat{Q}_3$ as:

$$\begin{aligned}
\hat{Q}_3 &= h^{1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q} Re\left\{\left[\hat{\phi}_t(u)-\phi_t(u)\right]\left[\hat{\phi}_0(u)-\phi_t(u)\right]^*\right\}W(u)du \\
&= h^{1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q} Re\left\{\left[\frac{1}{Th}\sum_{s=1}^{T}K\left(\frac{s-t}{Th}\right)\varepsilon_s(u)\right]\left[\frac{1}{T}\sum_{s=1}^{T}\varepsilon_s(u)\right]^*\right\}W(u)du \\
&= T^{-2}h^{-1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q}\sum_{s,r}^{T}K\left(\frac{s-t}{Th}\right)Re[\varepsilon_s(u)\varepsilon_r(u)^*]W(u)du \\
&= T^{-2}h^{-1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q}K(0)|\varepsilon_t(u)|^2W(u)du \\
&\quad +T^{-2}h^{-1/2}\sum_{t\neq r}^{T}\int_{\mathbb{R}^q}K(0)Re[\varepsilon_t(u)\varepsilon_r(u)^*]W(u)du \\
&\quad +T^{-2}h^{-1/2}\sum_{t\neq s}^{T}\int_{\mathbb{R}^q}K\left(\frac{s-t}{Th}\right)Re[\varepsilon_s(u)\varepsilon_t(u)^*]W(u)du \\
&\quad +T^{-2}h^{-1/2}\sum_{t\neq s}^{T}\int_{\mathbb{R}^q}K\left(\frac{s-t}{Th}\right)|\varepsilon_s(u)|^2W(u)du
\end{aligned}$$

$$
\begin{aligned}
& +T^{-2}h^{-1/2}\sum_{t\neq s\neq r}^{T}\int_{\mathbb{R}^q}K\left(\frac{s-t}{Th}\right)Re[\varepsilon_s(u)\varepsilon_r(u)^*]W(u)du \\
= \;& \hat{Q}_{31}+\hat{Q}_{32}+\hat{Q}_{33}+\hat{Q}_{34}+\hat{Q}_{35}.
\end{aligned}
$$

Next, we want to show that $\hat{Q}_{3j}=op(1)$ for $j=1,...,5$.

For $\hat{Q}_{31}$, it is straightforward to see that $\hat{Q}_{31}=K(0)/(Th)\hat{Q}_{21}=op(1)$ since $\hat{Q}_{21}=op(1)$. Similarly, $\hat{Q}_{32}=K(0)/(Th)\hat{Q}_{22}=op(1)$ given $\hat{Q}_{22}=op(1)$.

For $\hat{Q}_{33}$:

$$
\begin{aligned}
E(\hat{Q}_{33}) &= T^{-2}h^{-1/2}\sum_{t\neq s}^{T}\int_{\mathbb{R}^q}K\left(\frac{s-t}{Th}\right)E\left(Re[\varepsilon_s(u)\varepsilon_t(u)^*]\right)W(u)du \\
&= 2T^{-2}h^{-1/2}\sum_{1\leq t<s\leq T}\int_{\mathbb{R}^q}K\left(\frac{s-t}{Th}\right)E\left(Re[\varepsilon_s(u)\varepsilon_t(u)^*]\right)W(u)du \\
&= 2T^{-1}h^{-1/2}\sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right)K\left(\frac{j}{Th}\right)\int_{\mathbb{R}^q}E\left(Re[\varepsilon_1(u)\varepsilon_{1+j}(u)^*]\right)W(u)du \\
&\leq 2T^{-1}h^{-1/2}\sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right)K\left(\frac{j}{Th}\right)\int_{\mathbb{R}^q}|E\left(Re[\varepsilon_1(u)\varepsilon_{1+j}(u)^*]\right)|W(u)du \\
&\leq 2T^{-1}h^{-1/2}K(0)\sum_{j=1}^{T-1}\left(1-\frac{j}{T}\right)\int_{\mathbb{R}^q}|E\left(Re[\varepsilon_1(u)\varepsilon_{1+j}(u)^*]\right)|W(u)du \\
&\leq 2T^{-1}h^{-1/2}K(0)\,C\sum_{j=1}^{\infty}\beta(j)^{\frac{\delta}{1+\delta}} \\
&= O(T^{-1}h^{-1/2})=o(1).
\end{aligned}
$$

Furthermore, by Lemma A(ii) in Hjellvik et al. (1998),

$$
\begin{aligned}
E(\hat{Q}_{33}^2) &= T^{-4}h^{-1}E\Bigg[\sum_{1\leq t<s\leq T}\Big\{K\left(\frac{s-t}{Th}\right)\int_{\mathbb{R}^q}Re[\varepsilon_s(u)\varepsilon_t(u)^*]W(u)du \\
&\quad +K\left(\frac{t-s}{Th}\right)\int_{\mathbb{R}^q}\left(Re[\varepsilon_t(u)\varepsilon_s(u)^*]\right)W(u)du\Big\}\Bigg]^2 \\
&\leq CT^{-2}h^{-1}\max_{1\leq j\leq T-1}E\Bigg\{\Big[K\left(\frac{j}{Th}\right)\int_{\mathbb{R}^q}Re[\varepsilon_1(u)\varepsilon_{1+j}(u)^*]W(u)du \\
&\quad +K\left(\frac{-j}{Th}\right)\int_{\mathbb{R}^q}Re[\varepsilon_{1+j}(u)\varepsilon_1(u)^*]W(u)du\Big]^{2(1+\delta)}\Bigg\}^{\frac{1}{1+\delta}}\sum_{j=1}^{\infty}j\beta(j)^{\frac{\delta}{1+\delta}}
\end{aligned}
$$

$$= \quad o(1).$$

By Chebyshev's inequality, $\hat{Q}_{33} = op(1)$.

For $\hat{Q}_{34}$:

$$\begin{aligned}
\hat{Q}_{34} &= T^{-2}h^{-1/2}\sum_{t\neq s}^{T}\int_{\mathbb{R}^q} K\left(\frac{s-t}{Th}\right)|\varepsilon_s(u)|^2 W(u)du \\
&= T^{-1}h^{1/2}\sum_{s=1,s\neq t}^{T}\int_{\mathbb{R}^q}|\varepsilon_s(u)|^2 W(u)du\left[\frac{1}{Th}\sum_{t=1}^{T}K\left(\frac{s}{Th}-\frac{t}{Th}\right)\right] \\
&= T^{-1}h^{1/2}\sum_{s=1}^{T}\int_{\mathbb{R}^q}|\varepsilon_s(u)|^2 W(u)du\int_0^{\infty}K\left(\frac{s}{Th}-\eta\right)d\eta + o(1).
\end{aligned}$$

Then it is straightforward to show $\hat{Q}_{34} = op(1)$ by analogous arguments as in the proof of $\hat{R}_{11} = op(1)$.

Finally, for $\hat{Q}_{35}$:

$$\begin{aligned}
\hat{Q}_{35} &= T^{-2}h^{-1/2}\sum_{t\neq s\neq r}^{T}\int_{\mathbb{R}^q} K\left(\frac{s-t}{Th}\right)Re[\varepsilon_s(u)\varepsilon_r(u)^*]W(u)du \\
&= T^{-1}h^{1/2}\sum_{s\neq r}^{T}\int_{\mathbb{R}^q}Re[\varepsilon_s(u)\varepsilon_r(u)^*]W(u)du\left[\frac{1}{Th}\sum_{t=1,t\neq s}^{T}K\left(\frac{s-t}{Th}\right)\right] \\
&= T^{-1}h^{1/2}\sum_{s\neq r}^{T}\int_{\mathbb{R}^q}Re[\varepsilon_s(u)\varepsilon_r(u)^*]W(u)du\int_0^{1}K\left(\frac{s}{Th}-\eta\right)d\eta + o(1).
\end{aligned}$$

First, we show its mean is $o(1)$:

$$\begin{aligned}
E(\hat{Q}_{35}) &= T^{-1}h^{1/2}\sum_{s\neq r}^{T}\int_{\mathbb{R}^q}E(Re[\varepsilon_s(u)\varepsilon_r(u)^*])W(u)du\int_0^{\infty}K\left(\frac{s}{Th}-\eta\right)d\eta + o(1) \\
&= T^{-1}h^{1/2}\sum_{s=1}^{T-1}\sum_{r=s+1}^{T}\int_{\mathbb{R}^q}E(Re[\varepsilon_s(u)\varepsilon_r(u)^*])W(u)du\int_0^{1}K\left(\frac{s}{Th}-\eta\right)d\eta + o(1) \\
&\leq T^{-1}h^{1/2}\sum_{s=1}^{T-1}\sum_{r=s+1}^{T}\int_{\mathbb{R}^q}|E(Re[\varepsilon_s(u)\varepsilon_r(u)^*])|W(u)du\int_0^{1}K\left(\frac{s}{Th}-\eta\right)d\eta + o(1) \\
&\leq T^{-1}h^{1/2}K(0)\sum_{s=1}^{T-1}\sum_{r=s+1}^{T}\int_{\mathbb{R}^q}|E(Re[\varepsilon_s(u)\varepsilon_r(u)^*])|W(u)du + o(1)
\end{aligned}$$

$$= \quad O(h^{1/2}) = o(1).$$

Then we show its second moment is $o(1)$:

$$\begin{aligned}
E(\hat{Q}_{35}^2) &= T^{-2}hE\left[\sum_{s\neq r}^{T}\int_{\mathbb{R}^q} Re[\varepsilon_s(u)\varepsilon_r(u)^*]W(u)du\int_0^1 K\left(\frac{s}{Th}-\eta\right)d\eta\right]^2 + o(1) \\
&= T^{-2}hE\Bigg[\sum_{1\leq s<r\leq T}^{T}\int_{\mathbb{R}^q} Re[\varepsilon_s(u)\varepsilon_r(u)^*]W(u)du\int_0^1 K\left(\frac{s}{Th}-\eta\right)d\eta \\
&\quad + \int_{\mathbb{R}^q} Re[\varepsilon_r(u)\varepsilon_s(u)^*]W(u)du\int_0^1 K\left(\frac{r}{Th}-\eta\right)d\eta\Bigg]^2 + o(1) \\
&= O(h) = o(1),
\end{aligned}$$

by Lemma A(ii) of Hjellvik et al. (1998).

Finally, we show $\hat{Q}_3 = op(1)$. Intuitively, $\hat{Q}_3$ is a higher order term since it is a product of parametric estimation and nonparametric estimation. The parametric estimation has a faster rate of convergence to 0. This will make $\hat{Q}_3$ to converge to 0 faster than $\hat{Q}_1$. □

Proof of Theorem 2.4.1. Under $\mathbb{H}_A : \phi_t(u) \neq \phi_0(u)$,

$$\begin{aligned}
Th^{1/2}\hat{Q} &= h^{1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q}\left|\hat{\phi}_t(u) - \hat{\phi}_0(u)\right|^2 W(u)du \\
&= h^{1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q}\left|\hat{\phi}_t(u) - \phi_0(u)\right|^2 W(u)du \\
&\quad +h^{1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q}\left|\hat{\phi}_0(u) - \phi_0(u)\right|^2 W(u)du \\
&\quad -2h^{1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q} Re\left\{\left[\hat{\phi}_t(u) - \phi_0(u)\right]\left[\hat{\phi}_0(u) - \phi_0(u)\right]^*\right\} W(u)du \\
&= \hat{Q}_4 + \hat{Q}_5 - \hat{Q}_6.
\end{aligned}$$

First decompose $\hat{Q}_4$. By triangle inequality,

$$\hat{Q}_4 \quad = \quad h^{1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q}\left|\hat{\phi}_t(u) - \phi_t(u) + \phi_t(u) - \phi_0(u)\right|^2 W(u)du$$

$$
\begin{aligned}
&\leq \; 2h^{1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q}\left|\hat{\phi}_t(u)-\phi_t(u)\right|^2 W(u)du + 2h^{1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q}|\phi_t(u)-\phi_0(u)|^2 W(u)du \\
&= \; \hat{Q}_{41}+\hat{Q}_{42}.
\end{aligned}
$$

Then we study $\hat{Q}_{41}$:

$$
\begin{aligned}
\hat{Q}_{41} &= 2h^{1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q}\left|\hat{\phi}_t(u)-\phi_t(u)\right|^2 W(u)du \\
&= 2h^{1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q}\left|\sum_{s=t-\lfloor Th\rfloor}^{t+\lfloor Th\rfloor} Y_s e^{iu'Z_s}H_{st}-\phi_t(u)\right|^2 W(u)du \\
&= 2h^{1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q}\left|\sum_{s=t-\lfloor Th\rfloor}^{t+\lfloor Th\rfloor} [\phi_s(u)+\varepsilon_s(u)]H_{st}-\phi_t(u)\right|^2 W(u)du \\
&= h^{1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q}\left|\sum_{s=t-\lfloor Th\rfloor}^{t+\lfloor Th\rfloor} \varepsilon_s(u)H_{st}+\sum_{s=t-\lfloor Th\rfloor}^{t+\lfloor Th\rfloor}\phi_s(u)H_{st}-\phi_t(u)\right|^2 W(u)du \\
&\leq 4h^{1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q}\left|\sum_{s=t-\lfloor Th\rfloor}^{t+\lfloor Th\rfloor} \varepsilon_s(u)H_{st}\right|^2 W(u)du+4h^{1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q}\left|\sum_{s=t-\lfloor Th\rfloor}^{t+\lfloor Th\rfloor}\phi_s(u)H_{st}-\phi_t(u)\right|^2 W(u)du \\
&= \hat{Q}_{411}+\hat{Q}_{412}.
\end{aligned}
$$

Here $\hat{Q}_{411} = 4\hat{Q}_1$. And by the results in Proposition A.1.1 and Proposition A.1.2, $\hat{Q}_{411} = B + U + op(1)$, where $B = Op(h^{-1/2})$. For $\hat{Q}_{412}$, by the Taylor expansion of $\phi_s(u) \equiv \phi(u, \frac{s}{T})$ around t/T, we have

$$
\phi_s(u) = \phi_t(u)+\phi_t'(u)\left(\frac{s-t}{T}\right)+\frac{1}{2}\phi_t''(u)\left(\frac{s-t}{T}\right)^2+o\left(\frac{s-t}{T}\right)^2.
$$

Therefore,

$$
\begin{aligned}
\hat{Q}_{412} &= 4h^{1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q}\left|\sum_{s=t-\lfloor Th\rfloor}^{t+\lfloor Th\rfloor}\phi_s(u)H_{st}-\phi_t(u)\right|^2 W(u)du \\
&= 4h^{1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q}\left|\sum_{s=t-\lfloor Th\rfloor}^{t+\lfloor Th\rfloor}\left[\phi_t'(u)\left(\frac{s-t}{T}\right)+\frac{\phi_t''(u)}{2}\left(\frac{s-t}{T}\right)^2+o\left(\frac{s-t}{T}\right)^2\right]H_{st}\right|^2 W(u)du \\
&= 4h^{1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q}\left|\frac{1}{Th}\sum_{s=t-\lfloor Th\rfloor}^{t+\lfloor Th\rfloor}\left[\phi_t'(u)\left(\frac{s-t}{T}\right)+\frac{\phi_t''(u)}{2}\left(\frac{s-t}{T}\right)^2+o\left(\frac{s-t}{T}\right)^2\right]K\left(\frac{s-t}{Th}\right)\right|^2 W(u)du \\
&\quad +op(1)
\end{aligned}
$$

$$= 4h^{1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q}\left|\frac{1}{Th}\sum_{j=-\lfloor Th\rfloor}^{\lfloor Th\rfloor}\left[h\phi_t'(u)\left(\frac{j}{Th}\right)+\frac{h^2}{2}\phi_t''(u)\left(\frac{j}{Th}\right)^2+o\left(\frac{j}{T}\right)^2\right]K\left(\frac{j}{Th}\right)\right|^2 W(u)du$$
$$+op(1),$$

where the second to last equality comes from equivalence kernels. Let $\eta = \frac{j}{Th}$, since $\lfloor Th\rfloor \le j \le \lfloor Th\rfloor$, then $-1 \le \eta \le 1$ and by the Riemann approximation of an integral

$$\hat{Q}_{412} = 4h^{1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q}\left|\phi_t'(u)\int_{-1}^{1}\eta hK(\eta)\,d\eta+\frac{1}{2}\phi_t''(u)\int_{-1}^{1}\eta^2h^2K(\eta)\,d\eta+o(h^2)+o(1)\right|^2 W(u)du$$
$$+op(1)$$
$$= O(Th^{9/2}),$$

where the last equality comes from the assumption on kernel function $K(\cdot)$. Therefore, we have $\hat{Q}_{41} = Op(h^{-1/2})$. For $\hat{Q}_{42}$:

$$\lim_{T\to\infty} T^{-1}h^{-1/2}\hat{Q}_{42} = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\int_{\mathbb{R}^q}|\phi_t(u)-\phi_0(u)|^2\,W(u)du$$
$$= O(1),$$

where we use the fact that $E(Y_t^4) < \infty$ and $\int_{R^q}\|u\|_q^4W(u)du < \infty$. Thus, $\hat{Q}_{42} = O(Th^{1/2})$.

Combining $\hat{Q}_{41}$ and $\hat{Q}_{42}$, we show $\hat{Q}_4 = Op(Th^{1/2})$.

Next,

$$\hat{Q}_5 = h^{1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q}\left|\hat{\phi}_0(u)-\phi_0(u)\right|^2 W(u)du$$
$$= Th^{1/2}\int_{\mathbb{R}^q}\left|\frac{1}{T}\sum_{t=1}^{T}Y_te^{iu'Z_t}-\phi_0(u)\right|^2 W(u)du$$
$$= Th^{1/2}\int_{\mathbb{R}^q}\left|\frac{1}{T}\sum_{t=1}^{T}[\phi_t(u)+\varepsilon_t(u)]-\phi_0(u)\right|^2 W(u)du$$
$$= Th^{1/2}\int_{\mathbb{R}^q}\left|\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t(u)+\frac{1}{T}\sum_{t=1}^{T}[\phi_t(u)-\phi_0(u)]\right|^2 W(u)du$$

$$
\begin{aligned}
&= Th^{1/2}\int_{\mathbb{R}^q}\left|\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t(u)\right|^2 W(u)du + Th^{1/2}\int_{\mathbb{R}^q}\left|\frac{1}{T}\sum_{t=1}^{T}[\phi_t(u)-\phi_0(u)]\right|^2 W(u)du \\
&\quad +2Th^{1/2}\int_{\mathbb{R}^q} Re\left[\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t(u)\frac{1}{T}\sum_{t=1}^{T}[\phi_t(u)-\phi_0(u)]^*\right] \\
&= \hat{Q}_{51}+\hat{Q}_{52}+\hat{Q}_{53}.
\end{aligned}
$$

It is easy to see that $\hat{Q}_{51} = Op(h^{1/2})$ and $\hat{Q}_{52} = Op(Th^{1/2})$. And by triangle inequality,

$$
\begin{aligned}
\hat{Q}_{53} &\leq Th^{1/2}h^{1/2}\int_{\mathbb{R}^q}\left|\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t(u)\right|^2 W(u)du + Th^{1/2}\int_{\mathbb{R}^q}\left|\frac{1}{T}\sum_{t=1}^{T}[\phi_t(u)-\phi_0(u)]\right|^2 W(u)du \\
&= Op(Th^{1/2}).
\end{aligned}
$$

Therefore, it follows that $\hat{Q}_t = Op(Th^{1/2})$.

Finally, for $\hat{Q}_6$, by the triangle inequality,

$$
\begin{aligned}
\hat{Q}_6 &= 2h^{1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q} Re\left\{\left[\hat{\phi}_t(u)-\phi_0(u)\right]\left[\hat{\phi}_0(u)-\phi_0(u)\right]^*\right\} W(u)du \\
&\leq h^{1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q}\left|\hat{\phi}_t(u)-\phi_0(u)\right|^2 W(u)du + h^{1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q}\left|\hat{\phi}_0(u)-\phi_0(u)\right|^2 W(u)du^{1/2} \\
&\leq Op(h^{-1/2}+Th^{1/2}) \\
&= Op(Th^{1/2}).
\end{aligned}
$$

Consequently, $Th^{1/2}\hat{Q} = Op(Th^{1/2})$. Since we have shown that $\hat{B} = Op(h^{-1/2})$ and $\hat{V} = Op(1)$,

$$
\widehat{SQ} = \frac{Th^{1/2}\hat{Q}-\hat{B}}{\sqrt{\hat{V}}} = Op(Th^{1/2}).
$$

Thus,

$$
Pr(\widehat{SQ} > M_t) \to 1,
$$

as $T \to \infty$ under $\mathbb{H}_A$ for any non-stochastic constants $\{M_T = o(Th^{1/2})\}$. $\square$

Proof of Theorem 2.4.2. Under $\mathbb{H}_{a1}: g_t(X_t) - g_0(X_t) = \kappa_T h_t(X_t)$:

$$\begin{aligned}
\phi_t(u) - \phi_0(u) &= \int_{\mathbb{R}^q} [g_t(x) - g_0(x)]\, e^{iu'z} dF_{X_t,Z_t}(x,z) \\
&= \kappa_T \int_{\mathbb{R}^q} h_t(x) e^{iu'z} dF_{X_t,Z_t}(x,z) \\
&\equiv \kappa_T \psi_t(u).
\end{aligned}$$

Like in the proof of Theorem 2.4.1, we decompose

$$\begin{aligned}
Th^{1/2}\hat{Q} &= h^{1/2} \sum_{t=1}^{T} \int_{\mathbb{R}^q} \left|\hat{\phi}_t(u) - \hat{\phi}_0(u)\right|^2 W(u)du \\
&= h^{1/2} \sum_{t=1}^{T} \int_{\mathbb{R}^q} \left|\hat{\phi}_t(u) - \phi_0(u)\right|^2 W(u)du \\
&\quad + h^{1/2} \sum_{t=1}^{T} \int_{\mathbb{R}^q} \left|\hat{\phi}_0(u) - \phi_0(u)\right|^2 W(u)du \\
&\quad - 2h^{1/2} \sum_{t=1}^{T} \int_{\mathbb{R}^q} Re\left\{\left[\hat{\phi}_t(u) - \phi_0(u)\right]\left[\hat{\phi}_0(u) - \phi_0(u)\right]^*\right\} W(u)du \\
&= \hat{Q}_4 + \hat{Q}_5 + \hat{Q}_6.
\end{aligned}$$

We first work on $\hat{Q}_5$ and $\hat{Q}_6$. By the following identity

$$Y_t e^{iu'Z_t} = \phi_t(u) + \varepsilon_t(u),$$

it follows

$$\begin{aligned}
\hat{Q}_5 &= h^{1/2} \sum_{t=1}^{T} \int_{\mathbb{R}^q} \left|\hat{\phi}_0(u) - \phi_0(u)\right|^2 W(u)du \\
&= h^{1/2} \sum_{t=1}^{T} \int_{\mathbb{R}^q} \left|\frac{1}{T}\sum_{t=1}^{T} Y_t e^{iu'Z_t} - \phi_0(u)\right|^2 W(u)du \\
&= h^{1/2} \sum_{t=1}^{T} \int_{\mathbb{R}^q} \left|\frac{1}{T}\sum_{t=1}^{T} [\phi_t(u) + \varepsilon_t(u)] - \phi_0(u)\right|^2 W(u)du \\
&= h^{1/2} \sum_{t=1}^{T} \int_{\mathbb{R}^q} \left|\frac{1}{T}\sum_{t=1}^{T} [\phi_t(u) - \phi_0(u) + \varepsilon_t(u)]\right|^2 W(u)du \\
&= h^{1/2} \sum_{t=1}^{T} \int_{\mathbb{R}^q} \left|\frac{1}{T}\sum_{t=1}^{T} [\kappa_T \psi_t(u) + \varepsilon_t(u)]\right|^2 W(u)du
\end{aligned}$$

$$
\begin{aligned}
&= h^{1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q}\left|\frac{1}{T}\sum_{t=1}^{T}\kappa_T\psi_t(u)\right|^2 W(u)du + h^{1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q}\left|\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t(u)\right|^2 W(u)du \\
&\quad +2h^{1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q} Re\left\{\left[\frac{1}{T}\sum_{t=1}^{T}\kappa_T\psi_t(u)\right]\left[\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t(u)\right]^*\right\}W(u)du \\
&= \hat{Q}_{51}+\hat{Q}_{52}+\hat{Q}_{53}.
\end{aligned}
$$

Given $\kappa_T = T^{-1/2}h^{-1/4}$, we can show

$$
\begin{aligned}
\hat{Q}_{51} &= h^{1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q}\left|\frac{1}{T}\sum_{t=1}^{T}\kappa_T\psi_t(u)\right|^2 W(u)du \\
&= \frac{1}{T}\sum_{t=1}^{T}\int_{\mathbb{R}^q}\left|\frac{1}{T}\sum_{t=1}^{T}\psi_t(u)\right|^2 W(u)du \\
&= \int_{\mathbb{R}^q}\left|\frac{1}{T}\sum_{t=1}^{T}\psi_t(u)\right|^2 W(u)du \\
&= \int_{\mathbb{R}^q}\left|\int_0^1\psi(u,\tau)d\tau\right|^2 W(u)du + o(1).
\end{aligned}
$$

$\hat{Q}_{52}$ is identical to $\hat{Q}_2$. Since we have shown that $\hat{Q}_2 = op(1)$, thus $\hat{Q}_{52} = op(1)$. Given $\psi_t(u)$ is non-stochastic and $\kappa_T = T^{-1/2}h^{-1/4}$,

$$
\begin{aligned}
\hat{Q}_{53} &= 2h^{1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q} Re\left\{\left[\frac{1}{T}\sum_{t=1}^{T}\kappa_T\psi_t(u)\right]\left[\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t(u)\right]^*\right\}W(u)du \\
&= 2T^{-1/2}h^{1/4}\sum_{t=1}^{T}\int_{\mathbb{R}^q} Re\left\{\left[\int_0^1\psi(u,\tau)d\tau\right]\left[\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t(u)\right]^*\right\}W(u)du + op(1) \\
&= 2T^{1/2}h^{1/4}\int_{\mathbb{R}^q} Re\left\{\left[\int_0^1\psi(u,\tau)d\tau\right]\left[\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t(u)\right]^*\right\}W(u)du + op(1).
\end{aligned}
$$

The order of magnitude of $\hat{Q}_{53}$ is solely determined by $\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t(u)$. It is trivial to show that $\hat{Q}_{53} = Op(h^{1/4}) = op(1)$. Therefore, we have shown that

$$
\hat{Q}_5 = \int_{\mathbb{R}^q}\left|\int_0^1\psi(u,\tau)d\tau\right|^2 W(u)du + op(1).
$$

Next, for $\hat{Q}_6$:

$$
\hat{Q}_6 = -2h^{1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q} Re\left\{\left[\hat{\phi}_t(u)-\phi_0(u)\right]\left[\hat{\phi}_0(u)-\phi_0(u)\right]^*\right\}W(u)du
$$

$$
\begin{aligned}
&= -2h^{1/2} \sum_{t=1}^{T} \int_{\mathbb{R}^q} Re\left\{ \left[\hat{\phi}_t(u) - \phi_0(u) \right] \left[\frac{1}{T} \sum_{t=1}^{T} Y_t e^{iu'Z_t} - \phi_0(u) \right]^* \right\} W(u)du \\
&= -2h^{1/2} \sum_{t=1}^{T} \int_{\mathbb{R}^q} Re\left\{ \left[\hat{\phi}_t(u) - \phi_0(u) \right] \left[\frac{1}{T} \sum_{t=1}^{T} \kappa_T \psi_t(u) + \varepsilon_t(u) \right]^* \right\} W(u)du \\
&= -2h^{1/2} \sum_{t=1}^{T} \int_{\mathbb{R}^q} Re\left\{ \left[\hat{\phi}_t(u) - \phi_0(u) \right] \left[\frac{1}{T} \sum_{t=1}^{T} \kappa_T \psi_t(u) \right]^* \right\} W(u)du \\
&\quad -2h^{1/2} \sum_{t=1}^{T} \int_{\mathbb{R}^q} Re\left\{ \left[\hat{\phi}_t(u) - \phi_0(u) \right] \left[\frac{1}{T} \sum_{t=1}^{T} \varepsilon_t(u) \right]^* \right\} W(u)du \\
&\equiv \hat{Q}_{61} + \hat{Q}_{62}.
\end{aligned}
$$

We leave $\hat{Q}_6$ for now and work on $\hat{Q}_4$:

$$
\begin{aligned}
\hat{Q}_4 &= h^{1/2} \sum_{t=1}^{T} \int_{\mathbb{R}^q} \left| \hat{\phi}_t(u) - \phi_t(u) + \phi_t(u) - \phi_0(u) \right|^2 W(u)du \\
&= h^{1/2} \sum_{t=1}^{T} \int_{\mathbb{R}^q} \left| \hat{\phi}_t(u) - \phi_t(u) \right|^2 W(u)du + h^{1/2} \sum_{t=1}^{T} \int_{\mathbb{R}^q} \left| \phi_t(u) - \phi_0(u) \right|^2 W(u)du \\
&\quad +2h^{1/2} \sum_{t=1}^{T} \int_{\mathbb{R}^q} Re\left\{ \left[\hat{\phi}_t(u) - \phi_t(u) \right] \left[\phi_t(u) - \phi_0(u) \right]^* \right\} W(u)du \\
&= h^{1/2} \sum_{t=1}^{T} \int_{\mathbb{R}^q} \left| \hat{\phi}_t(u) - \phi_t(u) \right|^2 W(u)du + h^{1/2} \sum_{t=1}^{T} \int_{\mathbb{R}^q} \kappa_T^2 \left| \psi_t(u) \right|^2 W(u)du \\
&\quad +2h^{1/2} \sum_{t=1}^{T} \int_{\mathbb{R}^q} Re\left\{ \left[\hat{\phi}_t(u) - \phi_t(u) \right] \kappa_T \psi_t(u)^* \right\} W(u)du \\
&= \hat{Q}_{41} + \hat{Q}_{44} + \hat{Q}_{45}.
\end{aligned}
$$

In Proof of Theorem 2, we have shown $\hat{Q}_{41} = B + op(1)$. Given $\kappa_T = T^{-1/2}h^{-1/4}$, it follows

$$
\begin{aligned}
\hat{Q}_{44} &= h^{1/2} \sum_{t=1}^{T} \int_{\mathbb{R}^q} \kappa_T^2 \left| \psi_t(u) \right|^2 W(u)du \\
&= \frac{1}{T} \sum_{t=1}^{T} \int_{\mathbb{R}^q} \left| \psi_t(u) \right|^2 W(u)du \\
&= \int_{\mathbb{R}^q} \int_0^1 |\psi(u, \eta)|^2 W(u) d\eta du + op(1).
\end{aligned}
$$

Next,

$$
\begin{aligned}
\hat{Q}_{45} &= 2h^{1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q} Re\left\{\left[\hat{\phi}_t(u)-\phi_t(u)\right]\kappa_T\psi_t(u)^*\right\}W(u)du \\
&= 2h^{1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q} Re\left\{\left[\hat{\phi}_t(u)-\phi_0(u)+\phi_0(u)-\phi_t(u)\right]\kappa_T\psi_t(u)^*\right\}W(u)du \\
&= 2h^{1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q} Re\left\{\left[\hat{\phi}_t(u)-\phi_0(u)\right]\kappa_T\psi_t(u)^*\right\}W(u)du \\
&\quad -2h^{1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q} Re\left\{\left[\phi_t(u)-\phi_0(u)\right]\kappa_T\psi_t(u)^*\right\}W(u)du \\
&= 2h^{1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q} Re\left\{\left[\hat{\phi}_t(u)-\phi_0(u)\right]\kappa_T\psi_t(u)^*\right\}W(u)du \\
&\quad -2h^{1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q} \left|\kappa_T\psi_t(u)\right|^2 W(u)du \\
&= 2h^{1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q} Re\left\{\left[\hat{\phi}_t(u)-\phi_0(u)\right]\kappa_T\psi_t(u)^*\right\}W(u)du \\
&\quad -2\frac{1}{T}\sum_{t=1}^{T}\int_{\mathbb{R}^q} \left|\psi_t(u)\right|^2 W(u)du \\
&= 2h^{1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q} Re\left\{\left[\hat{\phi}_t(u)-\phi_0(u)\right]\kappa_T\psi_t(u)^*\right\}W(u)du \\
&\quad -2\int_{\mathbb{R}^q}\int_0^1 |\psi(u,\eta)|^2 W(u)d\eta du + op(1).
\end{aligned}
$$

We claim that $2h^{1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q} Re\left\{\left[\hat{\phi}_t(u)-\phi_0(u)\right]\kappa_T\psi_t(u)^*\right\}W(u)du + \hat{Q}_{61} = op(1)$:

$$
\begin{aligned}
& 2h^{1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q} Re\left\{\left[\hat{\phi}_t(u)-\phi_0(u)\right]\kappa_T\psi_t(u)^*\right\}W(u)du + \hat{Q}_{61} \\
&= 2h^{1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q} Re\left\{\left[\sum_{s=t-\lfloor Th\rfloor}^{t+\lfloor Th\rfloor}(\phi_s(u)+\varepsilon_s(u)-\phi_0(u))H_{st}\right]\left[\kappa_T\psi_t(u)-\frac{1}{T}\sum_{t=1}^{T}\kappa_T\psi_t(u)\right]^*\right\}W(u)du \\
&= 2h^{1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q} Re\left\{\left[\sum_{s=t-\lfloor Th\rfloor}^{t+\lfloor Th\rfloor}(\kappa_T\psi_s(u)+\varepsilon_s(u))H_{st}\right]\left[\kappa_T\psi_t(u)-\frac{1}{T}\sum_{t=1}^{T}\kappa_T\psi_t(u)\right]^*\right\}W(u)du \\
&= 2h^{1/2}\kappa_T^2\sum_{t=1}^{T}\int_{\mathbb{R}^q} Re\left\{\left[\sum_{s=t-\lfloor Th\rfloor}^{t+\lfloor Th\rfloor}\psi_s(u)H_{st}\right]\left[\psi_t(u)-\frac{1}{T}\sum_{t=1}^{T}\psi_t(u)\right]^*\right\}W(u)du
\end{aligned}
$$

$$+2h^{1/2}\kappa_T \sum_{t=1}^{T} \int_{\mathbb{R}^q} Re\left\{\left[\sum_{s=t-\lfloor Th\rfloor}^{t+\lfloor Th\rfloor} \varepsilon_s(u)H_{st}\right]\left[\psi_t(u) - \frac{1}{T}\sum_{t=1}^{T}\psi_t(u)\right]^*\right\} W(u)du$$
$$\equiv \quad R_8 + R_9.$$

By the weighted Cauchy Swartz inequality,

$$\begin{aligned}
R_8 &\leq 2h^{1/2}\kappa_T^2 \int_{\mathbb{R}^q} \left\{\left[\sum_{t=1}^{T}\left|\sum_{s=t-\lfloor Th\rfloor}^{t+\lfloor Th\rfloor}\psi_s(u)H_{st}\right|^2\right]^{1/2}\left[\sum_{t=1}^{T}\left|\psi_t(u) - \frac{1}{T}\sum_{t=1}^{T}\psi_t(u)\right|^2\right]^{1/2}\right\} W(u)du \\
&\leq 2\left[h^{1/2}\kappa_T^2 \sum_{t=1}^{T}\int_{\mathbb{R}^q}\left|\sum_{s=t-\lfloor Th\rfloor}^{t+\lfloor Th\rfloor}\psi_s(u)H_{st}\right|^2 W(u)du\right]^{1/2} \\
&\quad \times\left[h^{1/2}\kappa_T^2 \sum_{t=1}^{T}\int_{\mathbb{R}^q}\left|\psi_t(u) - \frac{1}{T}\sum_{t=1}^{T}\psi_t(u)\right|^2 W(u)du\right]^{1/2} \\
&= O(1) * O(T^{-1/2}) = o(1),
\end{aligned}$$

where the last equality needs $|\psi_s(u)|^2 < C$ for all $u \in \mathbb{R}^q$. And by the triangle inequality

$$\begin{aligned}
R_9 &\leq h^{1/2}\kappa_T \sum_{t=1}^{T}\int_{\mathbb{R}^q}\left|\sum_{s=t-\lfloor Th\rfloor}^{t+\lfloor Th\rfloor}\varepsilon_s(u)H_{st}\right|^2 W(u)du \\
&\quad +h^{1/2}\kappa_T \sum_{t=1}^{T}\int_{\mathbb{R}^q}\left|\psi_t(u) - \frac{1}{T}\sum_{t=1}^{T}\psi_t(u)\right|^2 W(u)du \\
&= Op(T^{-1/2}h^{-3/4}) + O(T^{-1/2}h^{1/4}) \\
&= op(1),
\end{aligned}$$

where we have used the result

$$h^{1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q}\left|\sum_{s=t-\lfloor Th\rfloor}^{t+\lfloor Th\rfloor}\varepsilon_s(u)H_{st}\right|^2 W(u)du = Op(h^{-1/2}),$$

obtained in Proposition 1. Therefore, we have shown that $R_8 + R_9 = op(1)$.

Lastly, we need to show $\hat{Q}_{62} = op(1)$:

$$\hat{Q}_{62} \quad = \quad -2h^{1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q} Re\left\{\left[\hat{\phi}_t(u) - \phi_0(u)\right]\left[\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t(u)\right]^*\right\} W(u)du$$

$$
\begin{aligned}
&= -2h^{1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q} Re\left\{\left[\sum_{s=t-\lfloor Th\rfloor}^{t+\lfloor Th\rfloor}(\kappa_T\psi_s(u)+\varepsilon_s(u))H_{st}\right]\left[\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t(u)\right]^*\right\}W(u)du \\
&= -2h^{1/2}\kappa_T\sum_{t=1}^{T}\int_{\mathbb{R}^q} Re\left\{\left[\sum_{s=t-\lfloor Th\rfloor}^{t+\lfloor Th\rfloor}\psi_s(u)H_{st}\right]\left[\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t(u)\right]^*\right\}W(u)du \\
&\quad -2h^{1/2}\sum_{t=1}^{T}\int_{\mathbb{R}^q} Re\left\{\left[\sum_{s=t-\lfloor Th\rfloor}^{t+\lfloor Th\rfloor}\varepsilon_s(u)H_{st}\right]\left[\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t(u)\right]^*\right\}W(u)du \\
&= \hat{Q}_{621}+\hat{Q}_{622}.
\end{aligned}
$$

Firstly, $\hat{Q}_{621} = Op(h^{1/4})$ since $\sum_{s=t-\lfloor Th\rfloor}^{t+\lfloor Th\rfloor}\psi_s(u)H_{st} = O(1)$ and $\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t(u) = Op(T^{-1/2})$. For $\hat{Q}_{622}$, we have shown it is $op(1)$ in the proof of Theorem A.1.3. Thus, $\hat{Q}_{62} = op(1)$.

Finally, we combine all the terms in $\hat{Q}_4$, $\hat{Q}_5$, and $\hat{Q}_6$:

$$
\begin{aligned}
Th^{1/2}\hat{Q} &= \hat{Q}_4+\hat{Q}_5+\hat{Q}_6 \\
&= B+\int_{\mathbb{R}^q}\left|\int_0^1\psi(u,\tau)d\tau\right|^2 W(u)du-\int_{\mathbb{R}^q}\int_0^1|\psi(u,\eta)|^2W(u)d\eta du+op(1).
\end{aligned}
$$

□

Proof of Equation (2.8) and (2.11). The proof for Eq.(2.8) is given in Hong, Wang and Wang (2014). Here we prove Eq.(2.11) in a more straightforward way. Let $z\in\mathbb{R}^q$ and $u\in\mathbb{R}^q$,

$$
\begin{aligned}
&\int_{\mathbb{R}^q} cos(u'z)W_L(u)du \\
&= \int_{\mathbb{R}^q} cos\left(u_1z_1+\sum_{j=2}^{q}u_jz_j\right)\prod_{j=1}^{q}\frac{\lambda_j}{2}e^{-\lambda_j|u_j|}du \\
&= \int_{\mathbb{R}^q}\left[cos(u_1z_1)cos\left(\sum_{j=2}^{q}u_jz_j\right)-sin(u_1z_1)sin\left(\sum_{j=2}^{q}u_jz_j\right)\right]\prod_{j=1}^{q}\frac{\lambda_j}{2}e^{-\lambda_j|u_j|}du \\
&= \int_{\mathbb{R}} cos(u_1z_1)\frac{\lambda_j}{2}e^{-\lambda_j|u_j|}du_1\int_{\mathbb{R}^{q-1}}cos\left(\sum_{j=2}^{q}u_jz_j\right)\left(\prod_{j=2}^{q}\frac{\lambda_j}{2}e^{-\lambda_j|u_j|}du_j\right) \\
&\quad -\int_{\mathbb{R}} sin(u_1z_1)\frac{\lambda_j}{2}e^{-\lambda_j|u_j|}du_1\int_{\mathbb{R}^{q-1}}sin\left(\sum_{j=2}^{q}u_jz_j\right)\left(\prod_{j=2}^{q}\frac{\lambda_j}{2}e^{-\lambda_j|u_j|}du_j\right)
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{\lambda_1^2}{\lambda_1^2+z_1^2}\int_{\mathbb{R}^{q-1}} cos\left(\sum_{j=2}^{q} u_j z_j\right)\left(\prod_{j=2}^{q}\frac{\lambda_j}{2}e^{-\lambda_j|u_j|}du_j\right)\\
&= \prod_{j=1}^{q}\frac{\lambda_j^2}{\lambda_j^2+z_j^2},
\end{aligned}
$$

where the second to last equality comes from the fact that $sin(\cdot)$ function is an odd function and $W_L(u)$ is a symmetric weighting function. The last equality is due to 3.893(2) of Jeffrey and Zwillinger (2007). □

Proof of Equation (2.10). Let $W_N(u) = \prod_{k=1}^{q}\frac{1}{\sqrt{2\pi\xi}}exp\left(\frac{u_k^2}{2\xi^2}\right)$, where ξ is the standard deviation of normal density which measures the dispersion of weightings assigned to each frequency u.

$$
\begin{aligned}
\hat{Q}_N &= \frac{1}{T}\sum_{t=1}^{T}\int_{\mathbb{R}^q}\left|\hat{\phi}_t(u)-\hat{\phi}_0(u)\right|^2 W_N(u)du\\
&= \frac{1}{T}\sum_{t=1}^{T}\int_{\mathbb{R}^q}\left|\hat{\phi}_t(u)\right|^2 W_N(u)du+\frac{1}{T}\sum_{t=1}^{T}\int_{\mathbb{R}^q}\left|\hat{\phi}_0(u)\right|^2 W_N(u)du\\
&\quad -\frac{2}{T}\sum_{t=1}^{T}\int_{\mathbb{R}^q} Re\left[\hat{\phi}_t(u)\hat{\phi}_0(u)^*\right]W_N(u)du\\
&= \frac{1}{T}\sum_{t=1}^{T}\int_{\mathbb{R}^q}\sum_{s,r\in\mathbb{S}_1} Y_sY_re^{iu(Z_s-Z_r)}H_{st}H_{rt}W_N(u)du+\frac{1}{T}\sum_{t=1}^{T}\int_{\mathbb{R}^q}\frac{1}{T^2}\sum_{s,r\in\mathbb{S}_2}Y_sY_re^{iu(Z_s-Z_r)}W_N(u)du\\
&\quad -\frac{2}{T}\sum_{t=1}^{T}\int_{\mathbb{R}^q}\frac{1}{T}\sum_{s\in\mathbb{S}_1,r\in\mathbb{S}_2}Re\left[Y_sY_re^{iu(Z_s-Z_r)}H_{st}\right]W_N(u)du\\
&= \frac{1}{T}\sum_{t=1}^{T}\left[\sum_{r,s\in\mathbb{S}_1}A_{sr}H_{st}H_{rt}+\frac{1}{T^2}\sum_{r,s\in\mathbb{S}_2}A_{sr}-\frac{2}{T}\sum_{s\in\mathbb{S}_1,r\in\mathbb{S}_2}A_{sr}H_{st}\right],
\end{aligned}
$$

where $A_{sr} = Y_sY_re^{-\frac{|Z_s-Z_r|_q^2\xi^2}{2}}$, $\mathbb{S}_1 = \{-\lfloor Th\rfloor, -\lfloor Th\rfloor+1, 2, ..., T+\lfloor Th\rfloor\}$, and $\mathbb{S}_2 = \{1, 2, ..., T\}$.

$$
\begin{aligned}
\hat{B}_N &= h^{-1/2}\int_{\mathbb{R}^q}|\hat{\Omega}(u,u)|W_N(u)du\int K^2(\eta)d\eta\\
&= h^{-1/2}\sum_{j=-p_T}^{p_T}k\left(\frac{j}{p_T}\right)\frac{1}{T}\sum_{t=1+|j|}^{T}\int_{\mathbb{R}^q}Re\left[\hat{\varepsilon}_t(u)\hat{\varepsilon}_{t-j}(u)^*\right]W_N(u)du\int K^2(\eta)d\eta.
\end{aligned}
$$

The key is to solve the integral using $W_N(u)$:

$$\begin{aligned}
& \int_{\mathbb{R}^q} Re\left[\hat{\varepsilon}_t(u)\hat{\varepsilon}_{t-j}(u)^*\right] W_N(u)du \\
= & \int_{\mathbb{R}^q} Re\left[(Y_t e^{iu'Z_t} - \hat{\phi}_t(u))(Y_{t-j}e^{iu'Z_{t-j}} - \hat{\phi}_{t-j}(u))^*\right] W_N(u)du \\
= & \int_{\mathbb{R}^q} Re\left[Y_t Y_{t-j} e^{iu'(Z_t - Z_{t-j})} - \hat{\phi}_t(u) Y_{t-j} e^{-iu'Z_{t-j}} - Y_t e^{iu'Z_t}\hat{\phi}_{t-j}(u)^* + \hat{\phi}_t(u)\hat{\phi}_{t-j}(u)^*\right] W_N(u)du \\
= & A_{t,t-j} - \sum_{s\in\mathbb{S}_1} A_{s,t-j}H_{st} - \sum_{s\in\mathbb{S}_1} A_{t,s-j}H_{st} + \sum_{s,r\in\mathbb{S}_1} A_{s,r-j}H_{st}H_{(r-j)(t-j)} \\
\equiv & \hat{\zeta}(t, t-j).
\end{aligned}$$

Consequently,

$$\begin{aligned}
\hat{B}_N &= h^{-1/2} \sum_{j=-p_T}^{p_T} k\left(\frac{j}{p_T}\right) \frac{1}{T} \sum_{t=1+|j|}^{T} \int_{\mathbb{R}^q} Re\left[\hat{\varepsilon}_t(u)\hat{\varepsilon}_{t-j}(u)^*\right] W_N(u)du \int K^2(\eta)d\eta \\
&= h^{-1/2} \sum_{j=-p_T}^{p_T} k\left(\frac{j}{p_T}\right) \frac{1}{T} \sum_{t=1+|j|}^{T} \hat{\zeta}(t, t-j) \int K^2(\eta)d\eta.
\end{aligned}$$

Lastly,

$$\begin{aligned}
\hat{V}_N &= 2\int_{\mathbb{R}^{2q}} |\hat{\Omega}(u,v)|^2 W_N(u)W_N(v)dudv \int \left[\int K(\eta)K(\eta+\lambda)d\eta\right]^2 d\lambda \\
&= 2\int_{\mathbb{R}^{2q}} \left|\sum_{j=-p_T}^{p_T} k\left(\frac{j}{p_T}\right) \frac{1}{T} \sum_{t=1+|j|}^{T} \hat{\varepsilon}_t(u)\hat{\varepsilon}_{t-j}(v)\right|^2 W_N(u)W_N(v)dudv \int \left[\int K(\eta)K(\eta+\lambda)d\eta\right]^2 d\lambda \\
&= 2\sum_{j,l=-p_T}^{p_T} k\left(\frac{j}{p_T}\right) k\left(\frac{l}{p_T}\right) \frac{1}{T^2} \sum_{s=1+|j|}^{T} \sum_{r=1+|l|}^{T} \int_{\mathbb{R}^q} Re[\hat{\varepsilon}_s(u)\hat{\varepsilon}_{s-j}(v)\hat{\varepsilon}_r(u)^*\hat{\varepsilon}_{r-l}(v)^*] \\
&\quad \times W_N(u)W_N(v)dudv \int \left[\int K(\eta)K(\eta+\lambda)d\eta\right]^2 d\lambda
\end{aligned}$$

The key is to solve the following integral using $W_N(u)$ and $W_N(v)$:

$$\begin{aligned}
& \int_{\mathbb{R}^q} Re[\hat{\varepsilon}_s(u)\hat{\varepsilon}_{s-j}(v)\hat{\varepsilon}_r(u)^*\hat{\varepsilon}_{r-l}(v)^*] W_N(u)W_N(v)dudv \\
= & \int_{\mathbb{R}^q} Re[\hat{\varepsilon}_s(u)\hat{\varepsilon}_r(u)^*]W_N(u)du \times \int_{\mathbb{R}^q} Re[\hat{\varepsilon}_{s-j}(v)\hat{\varepsilon}_{r-l}(v)^*]W_N(v)dv \\
= & \hat{\zeta}(s,r)\hat{\zeta}(s-j, r-l),
\end{aligned}$$

by the result obtained in solving $\hat{B}_N$. Consequently,

$$\hat{V}_N = 2\sum_{j,l=-p_T}^{p_T} k\left(\frac{j}{p_T}\right) k\left(\frac{l}{p_T}\right) \frac{1}{T^2} \sum_{s=1+|j|}^{T} \sum_{r=1+|l|}^{T} \hat{\zeta}(s,r)\hat{\zeta}(s-j, r-l)$$

$$\times \int \left[\int K(\eta) K(\eta + \lambda) d\eta \right]^2 d\lambda.$$

$\square$

Proof of Equation (2.13). The proof follows directly from the derivation in the proof of Equation (10) and proof of Equation (11). The only change is that

$$A_{sr} = Y_s Y_r \prod_{k=1}^{q} \frac{\lambda_k^2}{\lambda_k^2 + (Z_{sk} - Z_{rk})^2}.$$

It follows that $\widehat{SQ}_L$ has a closed-form expression by using $W_L(u)$. $\square$

APPENDIX B

APPENDIX OF CHAPTER 3

B.1 Figures and Tables

Table B.1: Empirical size with exogenous covariates ($T = 100$)

X_t		IID error		ARCH error		Heteroskedastic Error	
		$\hat{K}$	$\hat{C}$	$\hat{K}$	$\hat{C}$	$\hat{K}$	$\hat{C}$
No strutural change	5%	0.058	0.048	0.044	0.042	0.042	0.044
	10%	0.112	0.106	0.096	0.098	0.104	0.106
Smooth change	5%	0.048	0.046	0.034	0.036	0.060	0.056
	10%	0.094	0.094	0.088	0.088	0.116	0.108
Abrupt break	5%	0.058	0.062	0.052	0.050	0.056	0.048
	10%	0.108	0.108	0.102	0.106	0.124	0.112

Note: (i) Sample size = 100; (ii) Number of replication = 1000; (iii) Number of resampling = 499; (iv) $\hat{K}$ is computed using a gird of $\mathcal{U} = [0.01, 0.02, ..., 1]$; (v) $\hat{C}$ is computed using the standard normal weighting function; (vi) The critical values are computed by the resampling method in Section 3.2.

Table B.2: Empirical size with exogenous covariates ($T = 300$)

X_t		IID error		ARCH error		Heteroskedastic Error	
		$\hat{K}$	$\hat{C}$	$\hat{K}$	$\hat{C}$	$\hat{K}$	$\hat{C}$
No structural change	5%	0.052	0.048	0.047	0.047	0.051	0.049
	10%	0.108	0.105	0.101	0.010	0.102	0.107
Smooth change	5%	0.050	0.049	0.048	0.046	0.052	0.056
	10%	0.100	0.104	0.092	0.094	0.106	0.108
Abrupt break	5%	0.058	0.052	0.052	0.054	0.052	0.050
	10%	0.104	0.103	0.105	0.102	0.115	0.110

Note: (i) Sample size = 300; (ii) Number of replication = 1000; (iii) Number of resampling = 499; (iv) $\hat{K}$ is computed using a gird of $\mathcal{U} = [0.01, 0.02, ..., 1]$; (v) $\hat{C}$ is computed using the standard normal weighting function; (vi) The critical values are computed by the resampling method in Section 3.2.

Table B.3: Empirical power with exogenous covariates ($T = 100$)

X_t		DGP P.1			DGP P.2			DGP P.3		
		$\hat{K}$	$\hat{C}$	$\hat{H}$	$\hat{K}$	$\hat{C}$	$\hat{H}$	$\hat{K}$	$\hat{C}$	$\hat{H}$
No structural change	5%	0.423	0.462	0.336	0.478	0.412	0.414	0.286	0.192	0.204
	10%	0.550	0.572	0.474	0.605	0.548	0.552	0.402	0.340	0.307
Smooth change	5%	0.974	0.992	—	0.913	0.846	—	0.477	0.516	—
	10%	0.992	0.997	—	0.961	0.946	—	0.644	0.704	—
Abrupt break	5%	0.927	0.934	—	0.661	0.610	—	0.305	0.312	—
	10%	0.960	0.960	—	0.768	0.736	—	0.442	0.436	—

Note: (i) I.I.D. Errors; (ii) Sample size = 100; Number of replication = 1000; Number of resampling = 499; (iii) $\hat{K}$ is computed using a gird of $\mathcal{U} = [0.01, 0.02, ..., 1]$; (iv) $\hat{C}$ is computed using the standard normal weighting function; (v) The critical values of $\hat{K}$ and $\hat{C}$ are computed by the resampling method in Section 3.2; (vi) $\hat{H}$ is the generalized Hausman's test statistic in Chen and Hong (2012) robust to conditional heteroskedasticity; (vii) The bandwidth h is set at $T^{-1/5}/\sqrt{12}$. Kernel function is the uniform kernel; (viii) The critical values are computed by the wild bootstrap provided in Chen and Hong (2012).

Table B.4: Empirical power with exogenous covariates ($T = 300$)

X_t		DGP P.1			DGP P.2			DGP P.3		
		$\hat{K}$	$\hat{C}$	$\hat{H}$	$\hat{K}$	$\hat{C}$	$\hat{H}$	$\hat{K}$	$\hat{C}$	$\hat{H}$
No structural change	5%	0.812	0.887	0.806	0.912	0.907	0.898	0.703	0.688	0.691
	10%	0.922	0.936	0.902	0.937	0.921	0.918	0.951	0.918	0.905
Smooth change	5%	0.981	0.992	—	0.968	0.946	—	0.728	0.756	—
	10%	0.998	1.000	—	0.995	0.998	—	0.865	0.848	—
Abrupt break	5%	0.968	0.984	—	0.862	0.828	—	0.615	0.623	—
	10%	1.000	1.000	—	0.927	0.912	—	0.882	0.886	—

Note: (i) I.I.D. Errors; (ii) Sample size = 300; Number of replication = 1000; Number of resampling = 499; (iii) $\hat{K}$ is computed using a gird of $\mathcal{U} = [0.01, 0.02, ..., 1]$; (iv) $\hat{C}$ is computed using the standard normal weighting function; (v) The critical values of $\hat{K}$ and $\hat{C}$ are computed by the resampling method in Section 3.2; (vi) $\hat{H}$ is the generalized Hausman's test statistic in Chen and Hong (2012) robust to conditional heteroskedasticity; (vii) The bandwidth h is set at $T^{-1/5}/\sqrt{12}$. Kernel function is the uniform kernel; (viii) The critical values are computed by the wild bootstrap provided in Chen and Hong (2012).

Table B.5: Empirical size with endogenous covariates ($T = 100$)

X_t		IID error		ARCH error		Heteroskedastic Error	
		$\hat{K}^{IV}$	$\hat{C}^{IV}$	$\hat{K}^{IV}$	$\hat{C}^{IV}$	$\hat{K}^{IV}$	$\hat{C}^{IV}$
No strutural change	5%	0.050	0.040	0.050	0.044	0.044	0.048
	10%	0.118	0.118	0.120	0.118	0.090	0.092
Smooth change	5%	0.046	0.042	0.040	0.048	0.044	0.044
	10%	0.104	0.104	0.110	0.110	0.094	0.088
Abrupt break	5%	0.056	0.052	0.064	0.070	0.050	0.052
	10%	0.108	0.118	0.138	0.144	0.096	0.100

Note: (i) Sample size = 100; (ii) Number of replication = 1000; (iii) Number of resampling = 499; (iv) $\hat{K}^{IV}$ is computed using a gird of $\mathcal{U} = [0.01, 0.02, ..., 1]$; (v) $\hat{C}^{IV}$ is computed using the standard normal weighting function; (vi) The critical values are computed by the resampling method in Section 3.3.

Table B.6: Empirical size with endogenous covariates ($T = 300$)

X_t		IID error		ARCH error		Heteroskedastic Error	
		$\hat{K}^{IV}$	$\hat{C}^{IV}$	$\hat{K}^{IV}$	$\hat{C}^{IV}$	$\hat{K}^{IV}$	$\hat{C}^{IV}$
No strutural change	5%	0.052	0.046	0.050	0.045	0.051	0.049
	10%	0.110	0.111	0.110	0.110	0.092	0.095
Smooth change	5%	0.046	0.047	0.041	0.048	0.047	0.044
	10%	0.102	0.105	0.107	0.105	0.094	0.092
Abrupt break	5%	0.052	0.050	0.057	0.058	0.051	0.052
	10%	0.107	0.112	0.118	0.123	0.102	0.101

Note: (i) Sample size = 300; (ii) Number of replication = 1000; (iii) Number of resampling = 499; (iv) $\hat{K}^{IV}$ is computed using a gird of $\mathcal{U} = [0.01, 0.02, ..., 1]$; (v) $\hat{C}^{IV}$ is computed using the standard normal weighting function; (vi) The critical values are computed by the resampling method in Section 3.3.

Table B.7: Empirical power with endogenous covariates ($T = 100$)

X_t		DGP P.1			DGP P.2			DGP P.3		
		$\hat{K}^{IV}$	$\hat{C}^{IV}$	$\widehat{CH}$	$\hat{K}^{IV}$	$\hat{C}^{IV}$	$\widehat{CH}$	$\hat{K}^{IV}$	$\hat{C}^{IV}$	$\widehat{CH}$
No structural change	5%	1.000	1.000	0.806	1.000	1.000	0.886	1.000	1.000	0.764
	10%	1.000	1.000	0.912	1.000	1.000	0.935	1.000	1.000	0.856
Smooth change	5%	1.000	1.000	0.725	0.994	0.994	0.881	0.990	0.994	0.857
	10%	1.000	1.000	0.837	0.994	0.996	0.958	0.994	0.996	0.918
Abrupt break	5%	0.984	0.988	0.786	0.990	0.994	0.826	0.912	0.924	0.786
	10%	0.996	0.996	0.825	0.994	0.996	0.927	0.976	0.968	0.892

Note: (i) I.I.D. Errors; (ii) Sample size = 100; Number of replication = 1000; Number of resampling = 499; (iii) $\hat{K}^{IV}$ is computed using a gird of $\mathcal{U} = [0.01, 0.02, ..., 1]$; (iv) $\hat{C}^{IV}$ is computed using the standard normal weighting function; (v) The critical values of $\hat{K}^{IV}$ and $\hat{C}^{IV}$ are computed by the resampling method in Section 3.3; (vi) $\widehat{CH}$ is the generalized Hausman's test statistic in Chen (2015); (vii) The bandwidth h is set at $T^{-1/5}/\sqrt{12}$ for both the structural function and the reduced form. Kernel function is the uniform kernel; (viii) The critical values are computed by the wild bootstrap provided in Chen (2015).

Table B.8: Empirical power with endogenous covariates (T = 300)

X_t		DGP P.1			DGP P.2			DGP P.3		
		$\hat{K}^{IV}$	$\hat{C}^{IV}$	$\widehat{CH}$	$\hat{K}^{IV}$	$\hat{C}^{IV}$	$\widehat{CH}$	$\hat{K}^{IV}$	$\hat{C}^{IV}$	$\widehat{CH}$
No structural change	5%	1.000	1.000	0.925	1.000	1.000	0.892	1.000	1.000	0.899
	10%	1.000	1.000	0.982	1.000	1.000	0.982	1.000	1.000	0.952
Smooth change	5%	1.000	1.000	0.994	0.998	1.000	1.000	0.998	0.997	0.982
	10%	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Abrupt break	5%	0.999	0.989	0.963	0.996	1.000	0.989	0.972	0.967	0.912
	10%	1.000	1.000	0.998	1.000	1.000	1.000	1.000	1.000	0.998

Note: (i) I.I.D. Errors; (ii) Sample size = 300; Number of replication = 1000; Number of resampling = 499; (iii) $\hat{K}^{IV}$ is computed using a gird of $\mathcal{U} = [0.01, 0.02, ..., 1]$; (iv) $\hat{C}^{IV}$ is computed using the standard normal weighting function; (v) The critical values of $\hat{K}^{IV}$ and $\hat{C}^{IV}$ are computed by the resampling method in Section 3.3; (vi) $\widehat{CH}$ is the generalized Hausman's test statistic in Chen (2015); (vii) The bandwidth h is set at $T^{-1/5}/\sqrt{12}$ for both the structural function and the reduced form. Kernel function is the uniform kernel; (viii) The critical values are computed by the wild bootstrap provided in Chen (2015).

B.2 Mathematical Proofs

Throughout the appendix, "$\xrightarrow{p}$", "$\xrightarrow{d}$", and "$\Rightarrow$"denote convergence in probability, convergence in distribution, and weak convergence respectively. We use the same notations as in the chapter and we denote C to be a generic bounded constant, A^* to be the conjugate transpose of $A \in \mathbb{C}$, and $Re(A)$ to be the real part of $A \in \mathbb{C}$.

Proof of Lemma 3.2.1. Let $B_t \equiv X_tX_t' - E(X_tX_t)$ be a $d \times d$ matrix with each entry being $b_{jkt} = X_{jt}X_{kt} - E(X_{jt}X_{kt})$, where X_{jt} and X_{kt} are the jth and kth element respectively in X_t for $1 \leq j, k \leq d$. Then

$$\sup_{u \in \mathbb{R}} \left\| \frac{1}{T} \sum_{t=1}^{T} X_tX_t' e^{iu2\pi t/T} - \frac{1}{T} \sum_{t=1}^{T} E(X_tX_t') e^{iu2\pi t/T} \right\|$$

$$
\begin{aligned}
&= \sup_{u\in\mathbb{R}} \left\| \frac{1}{T}\sum_{t=1}^{T} B_t e^{iu2\pi t/T} \right\| \\
&= \sup_{u\in\mathbb{R}} \left(\sum_{j,k=1}^{d} \left| \frac{1}{T}\sum_{t=1}^{T} b_{jkt} e^{iu2\pi t/T} \right|^2 \right)^{1/2} \\
&= \sup_{u\in\mathbb{R}} \left[\sum_{j,k=1}^{d} \left(\frac{1}{T^2}\sum_{t=1}^{T} b_{jkt}^2 + \frac{1}{T^2}\sum_{s\neq t}^{T} b_{jkt} b_{jks} e^{iu2\pi(t-s)/T} \right) \right]^{1/2} \\
&\equiv R_1, \mathit{say}.
\end{aligned}
$$

Next, we show that $R_1 = op(1)$:

$$
\begin{aligned}
ER_1 &= E\left\{ \sup_{u\in\mathbb{R}} \left[\sum_{j,k=1}^{d} \left(\frac{1}{T^2}\sum_{t=1}^{T} b_{jkt}^2 + \frac{1}{T^2}\sum_{s\neq t}^{T} b_{jkt} b_{jks} e^{iu2\pi(t-s)/T} \right) \right]^{1/2} \right\} \\
&\leq \sup_{u\in\mathbb{R}} E\left\{ \left[\sum_{j,k=1}^{d} \left(\frac{1}{T^2}\sum_{t=1}^{T} b_{jkt}^2 + \frac{1}{T^2}\sum_{s\neq t}^{T} b_{jkt} b_{jks} e^{iu2\pi(t-s)/T} \right) \right]^{1/2} \right\} \\
&\leq \sup_{u\in\mathbb{R}} \left\{ E\left[\sum_{j,k=1}^{d} \left(\frac{1}{T^2}\sum_{t=1}^{T} b_{jkt}^2 + \frac{1}{T^2}\sum_{s\neq t}^{T} b_{jkt} b_{jks} e^{iu2\pi(t-s)/T} \right) \right] \right\}^{1/2},
\end{aligned}
$$

by Jensen's inequality. It is straightforward to show

$$
\begin{aligned}
& \sup_{u\in\mathbb{R}} \left\{ E\left[\sum_{j,k=1}^{d} \left(\frac{1}{T^2}\sum_{t=1}^{T} b_{jkt}^2 + \frac{1}{T^2}\sum_{s\neq t}^{T} b_{jkt} b_{jks} e^{iu2\pi(t-s)/T} \right) \right] \right\}^{1/2} \\
&= \sup_{u\in\mathbb{R}} \left\{ \sum_{j,k=1}^{d} \left(\frac{1}{T^2}\sum_{t=1}^{T} E(b_{jkt}^2) + \frac{1}{T^2}\sum_{s\neq t}^{T} E(b_{jkt} b_{jks}) e^{iu2\pi(t-s)/T} \right) \right\}^{1/2} \\
&= \sup_{u\in\mathbb{R}} \left\{ \sum_{j,k=1}^{d} \left(\frac{1}{T^2}\sum_{t=1}^{T} E(b_{jkt}^2) + \frac{2}{T^2}\sum_{1\leq s\leq t\leq T}^{T} E(b_{jkt} b_{jks}) e^{iu2\pi(t-s)/T} \right) \right\}^{1/2} \\
&\leq \sup_{u\in\mathbb{R}} \left\{ \sum_{j,k=1}^{d} \left(\frac{1}{T^2}\sum_{t=1}^{T} E(|X_{jt}X_{kt}|^2) \right. \right. \\
&\quad \left. \left. + \frac{2}{T^2}\sum_{1\leq s\leq t\leq T}^{T} \alpha(|s-t|)^{\frac{\delta-1}{\delta}} (E|X_{jt}X_{kt}|^{2\delta})^{\frac{1}{2\delta}} (E|X_{js}X_{ks}|^{2\delta})^{\frac{1}{2\delta}} e^{iu2\pi(t-s)/T} \right) \right\}^{1/2} \\
&\leq \sup_{u\in\mathbb{R}} \left\{ d\sum_{j=1}^{d} \left[\frac{1}{T^2}\sum_{t=1}^{T} E(|X_{jt}|^4 + \frac{1}{T}\sum_{j=1}^{T-1} \left(1-\frac{j}{T}\right) \alpha(|j|)^{\frac{\delta-1}{\delta}} (E|X_{jt}|^{4\delta})^{\frac{1}{\delta}} e^{iu2\pi j/T} \right] \right\}^{1/2}
\end{aligned}
$$

$$= \quad O(T^{-1/2}),$$

given $E|X_{jt}|^{4\delta} \leq C < \infty$ and $\sum_{j=1}^{\infty} \alpha(j)^{\frac{\delta-1}{\delta}} < C$.

Then, we can obtain

$$\begin{aligned} E(R_1^2) &= E\left\{\sup_{u\in\mathbb{R}}\left[\sum_{j,k=1}^{d}\left(\frac{1}{T^2}\sum_{t=1}^{T}b_{jkt}^2 + \frac{1}{T^2}\sum_{s\neq t}^{T}b_{jkt}b_{jks}e^{iu2\pi(t-s)/T}\right)\right]\right\} \\ &= \sup_{u\in\mathbb{R}}\left\{E\left[\sum_{j,k=1}^{d}\left(\frac{1}{T^2}\sum_{t=1}^{T}b_{jkt}^2 + \frac{1}{T^2}\sum_{s\neq t}^{T}b_{jkt}b_{jks}e^{iu2\pi(t-s)/T}\right)\right]\right\} \\ &= Op(T^{-1}). \end{aligned}$$

Thus, by Chebyshev's inequality, $R_1 = op(1)$. We then show $\sup_{u\in\mathbb{R}} \left\|\frac{1}{T}\sum_{t=1}^{T} B_t e^{iu2\pi t/T}\right\| = op(1)$. By similar derivation, given $E(|\varepsilon_t|^{4\delta}) < C < \infty$, we can show

$$\sup_{u_1,u_2\in\mathbb{R}^2}\left\|\hat{V}_{xx}(u_1,u_2) - V_{xx}(u_1,u_2)\right\| \xrightarrow{p} 0,$$

since we can view $u_1, u_2 \in \mathbb{R}^2$ as $v \in \mathbb{R}$ where $v \equiv u_1 - u_2$. □

Proof of Theorem 3.2.1. Given the definition of the WLS estimator and under $\mathbb{H}_0$

$$\begin{aligned} \sqrt{T}\left[\hat{\beta}(u) - \beta_0\right] &= \hat{Q}_{xx}^{-1}(u)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}X_t(u)\varepsilon_t \\ &= \hat{Q}_{xx}^{-1}(u)\hat{S}_T(u). \end{aligned}$$

We first prove that $\sqrt{T}\left[\hat{\beta}(u) - \beta_0\right]$ converges in distribution to a normal distribution for each fixed $u \in \mathbb{R}$. Then we show that $\hat{S}_T(u)$ is stochastically equicontinuous and establish the weak convergence result.

Let

$$U_T(u) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\tau' X_t(u)\varepsilon_t$$

$$= \sum_{t=1}^{T} \left[\frac{1}{\sqrt{T}} \sum_{j=1}^{d} \tau_j X_{jt}(u) \varepsilon_t \right],$$

where $\tau \in \mathbb{R}^d$ is any nonzero vector such that $\tau'\tau = 1$, and τ_j and $X_{jt}(u)$ represent the jth entry of $d \times 1$ column vectors τ and $X_t(u)$ respectively.

Furthermore we define

$$\begin{aligned}
\zeta_T^2(u) &\equiv E[U_T(u)^2] \\
&= \frac{1}{T} E\left[\sum_{s,t=1}^{T} \sum_{j,k=1}^{d} \tau_j X_{jt}(u) \varepsilon_t \tau_k X_{ks}(u)^* \varepsilon_s \right] \\
&= \frac{1}{T} \sum_{j,k=1}^{d} \tau_j \tau_k \left[\sum_{s,t=1}^{T} E[X_{jt}(u) \varepsilon_t X_{ks}(u)^* \varepsilon_s] \right] \\
&= \frac{1}{T} \sum_{j,k=1}^{d} \tau_j \tau_k \left[\sum_{t=1}^{T} E[X_{jt}(u) X_{kt}(u)^* \varepsilon_t^2] \right] \\
&= O(1).
\end{aligned}$$

for all $u \in \mathbb{R}$, where the second to last equality comes from the fact that $E(\varepsilon_t | I_{t-1}) = 0$ almost surely, and last equality is due to the fact that $E|X_{tj}|^{4\delta} < \infty$, $E|\varepsilon_t|^{4\delta} < \infty$ and $\tau_j^2 < 1$ for all $j = 1, 2, ..., d$.

By Cramer-wold device, to show joint normality, we just need to show

$$\zeta_T(u)^{-1} U_T(u) \xrightarrow{d} N(0, 1),$$

for each fixed $u \in \mathbb{R}$ and all τ.

Let

$$\Psi_t(u) \equiv \frac{1}{\sqrt{T}} \sum_{j=1}^{d} \tau_j X_{jt}(u) \varepsilon_t,$$

by Theorem 1.1 in Bradley and Tone (2015), we need to show the following:

(i) $\alpha'(\{\Psi_t(u)\}_{t=1}^{T}, m) \to 0$, as $m \to \infty$;

(ii) $\rho'(\{\Psi_t(u)\}_{t=1}^{T}, 1) < 1$;

(iii) $\zeta_T^2(u) > 0$;

(iv) $E(\Psi_t(u)^2) = \sum_{j,k=1}^d \tau_j \tau_k E(X_{jt} X_{kt} \varepsilon_t^2) < \infty$;

(v) The Lindeberg condition:

$$\lim_{T\to\infty} \frac{1}{\zeta_T^2(u)} \sum_{t=1}^T E\left[\Psi_t(u)^2 \mathbf{1}\left(|\Psi_t(u)| > \zeta_T(u)\epsilon\right)\right] = 0, \ \forall \epsilon > 0.$$

Firstly, condition (i) can be implied by Assumption 3.2.1.

$$\begin{aligned} \alpha'(\{\Psi_t(u)\}_{t=1}^T, m) &= \sup_{1\le t\le T-m} \alpha\left[\sigma\left(\sum_{j=1}^d \tau_j X_{jt}(u)\varepsilon_t\right), \sigma\left(\sum_{j=1}^d \tau_j X_{jt+m}(u)\varepsilon_{t+m}\right)\right] \\ &= \sup_{1\le t\le T-m} \alpha[\sigma(X_t\varepsilon_t), \sigma(X_{t+m})\varepsilon_{t+m}] \\ &\le Cm^{-\nu}, \end{aligned}$$

for all u. Thus $\alpha'(\{\Psi_t(u)\}_{t=1}^T, m) \to 0$ as $m \to \infty$.

For condition (ii), notice that $\Psi_t(u)$ is martingale difference sequence process:

$$E(\Psi_t(u)|I_{t-1}) = E\left(\frac{1}{\sqrt{T}} \sum_{j=1}^d \tau_j X_{jt}(u)\varepsilon_t \middle| I_{t-1}\right) = 0,$$

almost surely. Thus

$$\rho'(\{\Psi_t(u)\}_{t=1}^T, 1) = 0,$$

where the maximal coefficient of correlation

$$\rho(\mathcal{A}, \mathcal{B}) \equiv \sup_{f\in L^2(\mathcal{A}), g\in L^2(\mathcal{B})} |corr(f, g)|,$$

and $L^2(\mathcal{A})$ consists any $\mathcal{A}$-measurable random variables with finite second moments.

Condition (iii) is satisfied since $\zeta_T^2(u) = O(1)$.

Condition (iv) is satisfied due to $E|X_{tj}|^{4\delta} < \infty$, $E|\varepsilon_t|^{4\delta} < \infty$ and $\tau_j^2 < 1$ for all $j = 1, 2, ..., d$.

At last, we verify the Lindeberg condition: $\forall\epsilon > 0$,

$$
\begin{aligned}
& \lim_{T\to\infty} \frac{1}{\zeta_T^2(u)} \sum_{t=1}^{T} E\left[\Psi_t(u)^2 \mathbf{1}\left(|\Psi_t(u)| > \zeta_T(u)\epsilon\right)\right] \\
\leq & \lim_{T\to\infty} \frac{1}{\zeta_T^2(u)} \sum_{j=1}^{d} \left(E\left[\Psi_t(u)^4\right]\right)^{1/2} E\left[\mathbf{1}\left(|\Psi_t(u)| > \zeta_T(u)\epsilon\right)\right]^{1/2} \\
\leq & \lim_{T\to\infty} \frac{1}{\zeta_T^2(u)} \sum_{j=1}^{d} \left(E\left[\Psi_t(u)^4\right]\right)^{1/2} \left(\frac{1}{\zeta_T^2(u)\epsilon^2} E\left[\Psi_t(u)^2\right]\right)^{1/2} \\
= & \lim_{T\to\infty} \frac{1}{\zeta_T^3(u)\epsilon} \sum_{j=1}^{d} \left(E\left[\Psi_t(u)^4\right]\right)^{1/2} \left(E\left[\Psi_t(u)^2\right]\right)^{1/2} \\
= & \lim_{T\to\infty} \frac{1}{\zeta_T^3(u)\epsilon} \sum_{j=1}^{d} \left(\frac{1}{T^2} E\left[\sum_{j=1}^{d} \tau_j X_{jt}(u)\right]^4\right)^{1/2} \left(\frac{1}{T} E\left[\sum_{j=1}^{d} \tau_j X_{jt}(u)\right]^2\right)^{1/2} \\
= & O(T^{-3/2}) \\
= & o(1),
\end{aligned}
$$

where the first inequality is due to the Cauchy-Schwartz inequality given $E[\Psi_t(u)^2] = E[|\Psi_t(u)^2|]$, $E[\mathbf{1}(|\Psi_t(u)| > \zeta_T(u)\epsilon)] = E[|\mathbf{1}(|\Psi_t(u)| > \zeta_T(u)\epsilon)|]$ due to the non-negativity of indicator function, and $E[\mathbf{1}\left(\left|\Psi_j(T,u)\right| > \zeta_T(u)\epsilon\right)^2] = E[\mathbf{1}\left(\left|\Psi_j(T,u)\right| > \zeta_T(u)\epsilon\right)]$. The second inequality is due to the Chebyshev's Inequality. Thus, we have verified condition (iv).

By Theorem 1.1 in Bradley and Tone (2015), for each $u \in \mathbb{R}$,

$$\zeta_T(u)^{-1} U_T(u) \xrightarrow{d} N(0,1).$$

Since this result holds for all $\tau'\tau = 1$, by Cramer-Wold Device,

$$S_T(u) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} X_t(u)\varepsilon_t \xrightarrow{d} N[0, V_{xx}(u,u)],$$

for each fixed $u \in \mathbb{R}$, where

$$V_{xx}(u_1,u_2) \;=\; \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} E[X_t(u_1) X_t(u_2)^{*'} \varepsilon_t^2].$$

Next, we show $S_T(u)$ is stochastically equicontinuous. By definition, we need show that for all $\epsilon > 0$ and $\eta > 0$, there exists $\vartheta > 0$ such that

$$\limsup_{T\to\infty} Pr\left(\sup_{u\in\mathbb{R}} \sup_{u'\in B(u,\vartheta)} \|S_T(u) - S_T(u')\| > \eta\right) < \epsilon,$$

where $B(u,\vartheta)$ is an open set in $\mathbb{R}$ such that $|u' - u| \leq \vartheta$ for all $u' \in B(u,\vartheta)$. Here the structure of $S_T(u)$ is quite straightforward and the Fourier basis function is a smooth function of $u \in \mathbb{R}$. So we can prove this condition directly.

By Taylor expansion, there exists an $\bar{u} \in B(u,\vartheta)$ lies in between u and u' such that

$$\begin{aligned} S_T(u) &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} X_t \varepsilon_t e^{iu2\pi t/T} \\ &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} X_t \varepsilon_t \left[e^{iu'2\pi t/T} + \frac{i2\pi t}{T} e^{i\bar{u}2\pi t/T} (u - u') \right], \end{aligned}$$

then we have

$$\|S_T(u) - S_T(u')\| = \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} X_t \varepsilon_t e^{i\bar{u}2\pi t/T} \frac{i2\pi t}{T} \right\| |u - u'|.$$

Thus

$$\begin{aligned} & \limsup_{T\to\infty} Pr\left(\sup_{u\in\mathbb{R}} \sup_{u'\in B(u,\vartheta)} \|S_T(u) - S_T(u')\| > \eta\right) \\ = & \limsup_{T\to\infty} Pr\left(\sup_{u\in\mathbb{R}} \sup_{u'\in B(u,\vartheta)} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} X_t \varepsilon_t e^{i\bar{u}2\pi t/T} \frac{i2\pi t}{T} \right\| |u - u'| > \eta\right) \\ \leq & \frac{1}{\eta^2} \limsup_{T\to\infty} E\left(\sup_{u\in\mathbb{R}} \sup_{u'\in B(u,\vartheta)} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} X_t \varepsilon_t e^{i\bar{u}2\pi t/T} \frac{i2\pi t}{T} \right\| |u - u'|\right)^2 \\ = & \frac{4\pi^2}{\eta^2} \limsup_{T\to\infty} E\left(\sup_{u\in\mathbb{R}} \sup_{u'\in B(u,\vartheta)} \sqrt{\sum_{j=1}^{d} \frac{1}{T} \sum_{s,t=1}^{T} X_{jt} X_{js} \varepsilon_t \varepsilon_s e^{i\bar{u}2\pi(t-s)/T} \frac{st}{T^2}} |u - u'|\right)^2 \\ \leq & \frac{4\pi^2}{\eta^2} \limsup_{T\to\infty} E\left(\sup_{u\in\mathbb{R}} \sup_{u'\in B(u,\vartheta)} \sum_{j=1}^{d} \frac{1}{T} \sum_{s,t=1}^{T} X_{jt} X_{js} \varepsilon_t \varepsilon_s e^{i\bar{u}2\pi(t-s)/T} \frac{st}{T^2} |u - u'|^2\right) \end{aligned}$$

$$\leq \frac{4\pi^2}{\eta}\vartheta^2 C_1,$$

where

$$C_1 \geq d \max_j E(\|X_t\|^4)E(\varepsilon_{jt}^4) < \infty.$$

Therefore, for any $\epsilon > 0$ and $\eta > 0$, we need to choose $\vartheta < \frac{\sqrt{\eta}\epsilon}{2\pi\sqrt{C_1}}$ such that

$$\frac{4\pi^2}{\eta}\vartheta^2 C_1 < \epsilon.$$

Therefore, $S_T(u)$ is stochastically equicontinuous over $u \in \mathbb{R} \setminus \mathbb{Z}$. Given Lemma 3.2.1, and by Slutsky theorem, we have

$$\sqrt{T}\left[\hat{\beta}(u) - \beta_0\right] \Rightarrow G(u),$$

where $G(u)$ is a zero-mean complex-valued Gaussian process with covariance kernel

$$\begin{aligned} K(u_1, u_2) &= cov\left[G(u_1), G(u_2)^*\right] \\ &= Q_{xx}^{-1}(u_1)V_{xx}(u_1, u_2)Q_{xx}^{-1}(u_2)^*. \end{aligned}$$

In particular, when $u = 0$,

$$\sqrt{T}\left[\hat{\beta}(0) - \beta_0\right] \xrightarrow{d} N[0, Q_{xx}^{-1}(0)V_{xx}(0,0)Q_{xx}^{-1}(0)].$$

Given Assumptions 3.2.1 to 3.2.5, under $\mathbb{H}_0$,

$$\begin{aligned} \sqrt{T}\hat{A}(u) &= \hat{Q}_{xx}(u)\sqrt{T}\left[\hat{\beta}(u) - \hat{\beta}(0)\right] \\ &= \hat{Q}_{xx}(u)\left\{\sqrt{T}\left[\hat{\beta}(u) - \beta_0\right] - \sqrt{T}\left[\hat{\beta}(0) - \beta_0\right]\right\} \\ &\Rightarrow Q_{xx}(u)[G(u) - G(0)] \\ &= \mathcal{G}(u), \end{aligned}$$

where $\mathcal{G}(u)$ is a complex-valued Gaussian process with mean 0 and variance covariance kernel

$$\mathcal{K}(u_1, u_2) = cov[\mathcal{G}(u_1), \mathcal{G}(u_2)^*]$$

$$
\begin{aligned}
&= cov\left\{Q_{xx}(u_1)[G(u_1)-G(0)],[G(u_2)^{*'}-G(0)']Q_{xx}(u_2)^*\right\}\\
&= Q_{xx}(u_1)cov\left[G(u_1),G(u_2)^{*'}\right]Q_{xx}(u_2)^*-Q_{xx}(u_1)cov\left[G(u_1),G(0)^{*'}\right]Q_{xx}(u_2)^*\\
&\quad -Q_{xx}(u_1)cov\left[G(0),G(u_2)^{*'}\right]Q_{xx}(u_2)^*+Q_{xx}(u_1)cov\left[G(0),G(0)^{*'}\right]Q_{xx}(u_2)^*\\
&= Q_{xx}(u_1)K(u_1,u_2)Q_{xx}(u_2)^*-Q_{xx}(u_1)K(u_1,0)Q_{xx}(u_2)^*\\
&\quad -Q_{xx}(u_1)K(0,u_2)Q_{xx}(u_2)^*+Q_{xx}(u_1)K(0,0)Q_{xx}(u_2)^*\\
&= Q_{xx}(u_1)\left[K(u_1,u_2)-K(0,u_2)-K(u_1,0)+K(0,0)\right]Q_{xx}(u_2)^*.
\end{aligned}
$$

□

Proof of Theorem 3.2.2. Under $\mathbb{H}_A$,

$$
\begin{aligned}
\hat{\beta}(u) &= \hat{Q}_{xx}^{-1}(u)\frac{1}{T}\sum_{t=1}^{T}X_t(u)Y_t\\
&= \hat{Q}_{xx}^{-1}(u)\frac{1}{T}\sum_{t=1}^{T}X_t(u)X_t'\beta_t+\hat{Q}_{xx}^{-1}(u)\frac{1}{T}\sum_{t=1}^{T}X_t(u)\varepsilon_t\\
&= \hat{Q}_{xx}^{-1}(u)\frac{1}{T}\sum_{t=1}^{T}X_t(u)X_t'\beta_t+op(1),
\end{aligned}
$$

where the last equality comes form the result in Theorem 3.1. We first show

$$
\sup_{u\in\mathbb{R}}\left\|\frac{1}{T}\sum_{t=1}^{T}X_t(u)X_t'\beta_t-\frac{1}{T}\sum_{t=1}^{T}E(X_t(u)X_t')\beta_t\right\|\xrightarrow{p}0,
$$

as $T\to\infty$. Following the same notation as in Proof of Lemma 3.1, we have

$$
\begin{aligned}
&\sup_{u\in\mathbb{R}}\left\|\frac{1}{T}\sum_{t=1}^{T}X_t(u)X_t'\beta_t-\frac{1}{T}\sum_{t=1}^{T}E(X_t(u)X_t')\beta_t\right\|\\
=\ &\sup_{u\in\mathbb{R}}\left\|\frac{1}{T}\sum_{t=1}^{T}B_t\beta_t e^{iu2\pi t/T}\right\|\\
=\ &\sup_{u\in\mathbb{R}}\left[\sum_{j=1}^{d}\left|\frac{1}{T}\sum_{t=1}^{T}\sum_{k=1}^{d}b_{jkt}\beta_{jt}e^{iu2\pi t/T}\right|^2\right]^{1/2}\\
=\ &\sup_{u\in\mathbb{R}}\left\{\sum_{j=1}^{d}\left[\frac{1}{T^2}\sum_{t=1}^{T}\left(\sum_{k=1}^{d}b_{jkt}\beta_{jt}\right)^2+\frac{1}{T^2}\sum_{s\neq t}^{T}\left(\sum_{k_1,k_2=1}^{d}b_{jk_1t}b_{jk_2s}\beta_{jt}\beta_{js}\right)e^{iu2\pi(t-s)/T}\right]\right\}^{1/2}
\end{aligned}
$$

$$\equiv \quad R_2, say.$$

Then

$$\begin{aligned}
ER_2 &\leq \sup_{u\in\mathbb{R}} E\left\{\sum_{j=1}^{d}\left[\frac{1}{T^2}\sum_{t=1}^{T}\left(\sum_{k=1}^{d} b_{jkt}\beta_{jt}\right)^2 + \frac{1}{T^2}\sum_{s\neq t}^{T}\left(\sum_{k_1,k_2=1}^{d} b_{jk_1t}b_{jk_2s}\beta_{jt}\beta_{js}\right)e^{iu2\pi(t-s)/T}\right]\right\}^{1/2}\\
&\leq \sup_{u\in\mathbb{R}}\left\{\sum_{j=1}^{d} E\left[\frac{1}{T^2}\sum_{t=1}^{T}\left(\sum_{k=1}^{d} b_{jkt}\beta_{jt}\right)^2 + \frac{1}{T^2}\sum_{s\neq t}^{T}\left(\sum_{k_1,k_2=1}^{d} b_{jk_1t}b_{jk_2s}\beta_{jt}\beta_{js}\right)e^{iu2\pi(t-s)/T}\right]\right\}^{1/2}\\
&= \sup_{u\in\mathbb{R}}\left\{\sum_{j=1}^{d}\left[\frac{1}{T^2}\sum_{t=1}^{T} E\left(\sum_{k=1}^{d} b_{jkt}\beta_{jt}\right)^2 + \frac{1}{T^2}\sum_{s\neq t}^{T} E\left(\sum_{k_1,k_2=1}^{d} b_{jk_1t}b_{jk_2s}\beta_{jt}\beta_{js}\right)e^{iu2\pi(t-s)/T}\right]\right\}^{1/2}\\
&= Op(T^{-1/2}),
\end{aligned}$$

given similar argument as proof in R_1 and the fact that $\sum_{t=1}^{T}\|\beta_t\|^2 < C < \infty$ We show that $R_2(u) = op(1)$. Similarly, we can show $ER_2^2 = Op(T^{-1})$ and thus $R_2 = op(1)$ by Chebyshev's inequality.

Under $\mathbb{H}_A$,

$$\begin{aligned}
\hat{A}(u) &= \hat{Q}_{xx}(u)[\hat{\beta}(u) - \hat{\beta}(0)]\\
&= \frac{1}{T}\sum_{t=1}^{T} X_t(u)Y_t - \hat{Q}_{xx}(u)\hat{\beta}(0)\\
&= \frac{1}{T}\sum_{t=1}^{T} X_t(u)X_t'\beta_t - \hat{Q}_{xx}(u)\hat{\beta}(0) + op(1)\\
&= \frac{1}{T}\sum_{t=1}^{T}[X_tX_t' - E(X_tX_t')]\beta_t e^{iu2\pi t/T} + \frac{1}{T}\sum_{t=1}^{T} E(X_tX_t')\beta_t e^{iu2\pi t/T}\\
&\quad -\left\{[\hat{Q}_{xx}(u) - Q_{xx}(u)]\hat{\beta}(0) + Q_{xx}(u)\hat{\beta}(0)\right\} + op(1)
\end{aligned}$$

then

$$\begin{aligned}
&\sup_{u\in\mathbb{R}}\left\|\hat{A}(u) - \frac{1}{T}\sum_{t=1}^{T}\left[I_d e^{iu2\pi t/T} - Q_{xx}(u)Q_{xx}^{-1}(0)\right]E(X_tX_t')\beta_t\right\|\\
= &\sup_{u\in\mathbb{R}}\left\|\hat{A}(u) - \frac{1}{T}\sum_{t=1}^{T} E(X_tX_t')\beta_t e^{iu2\pi t/T} + Q_{xx}(u)Q_{xx}^{-1}(0)\sum_{t=1}^{T} E(X_tX_t')\beta_t\right\|
\end{aligned}$$

$$
\begin{aligned}
&= \sup_{u\in\mathbb{R}} \left\| \frac{1}{T}\sum_{t=1}^{T}[X_tX_t' - E(X_tX_t')]\beta_t e^{iu2\pi t/T} - [\hat{Q}_{xx}(u) - Q_{xx}(u)]\hat{\beta}(0) \right. \\
&\quad \left. -Q_{xx}(u)\left[\hat{\beta}(0) - Q_{xx}^{-1}(0)\frac{1}{T}\sum_{t=1}^{T}E(X_tX_t')\beta_t\right] \right\| \\
&\leq \sup_{u\in\mathbb{R}} \left\| \frac{1}{T}\sum_{t=1}^{T}[X_tX_t' - E(X_tX_t')]\beta_t e^{iu2\pi t/T} \right\| \\
&\quad + \sup_{u\in\mathbb{R}} \left\| [\hat{Q}_{xx}(u) - Q_{xx}(u)]\hat{\beta}(0) \right\| + \sup_{u\in\mathbb{R}} \left\| Q_{xx}(u)\left[\hat{\beta}(0) - Q_{xx}^{-1}(0)\frac{1}{T}\sum_{t=1}^{T}E(X_tX_t')\beta_t\right] \right\| \\
&= R_3 + R_4 + R_5,
\end{aligned}
$$

By triangle inequality. First, it is obvious to see that $R_3 = R_2 = op(1)$. We also have $R_4 = op(1)$ given Lemma 3.2.1. And finally $R_5 = op(1)$ is a standard result from OLS.

□

Proof of Theorem 3.2.3. Theorem 3.2.1 established the weak convergence of $\sqrt{T}\hat{A}(u)$ to a complex-valued Gaussian process $\mathcal{G}(u)$. By continuous mapping theorem, we have

$$\hat{K} \xrightarrow{d} \sup_{u\in\mathbb{R}} \|\mathcal{G}(u)\|^2,$$

and

$$\hat{C} \xrightarrow{d} \int_{\mathbb{R}} \|\mathcal{G}(u)\|^2 W(u)du.$$

□

Proof of Corollary 3.2.1. We refer to Theorem 2 in Hansen (1996) by showing that Assumption 1 and Assumption 2 are satisfied.

Firstly, $\{X_t, Y_t\}_{t=1}^T$ is an absolutely regular process under Assumption 3.2.1. The stationarity assumption in Hansen (1996) is equivalent to the uniform boundedness of $E(|X_{jt}|)$ for all j and all t. The regression score in this chapter is

$$s_T(u) \equiv \hat{M}_t(u)\varepsilon_t.$$

Therefore we need to show

$$\inf_{u\in\mathbb{R}}\left\{\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}E[\hat{M}_t(u)\hat{M}_t(u)^*]\right\} > 0.$$

Given $\hat{M}_t(u) = X_t e^{iu2\pi t/T} - \hat{Q}_{xx}(u)\hat{Q}_{xx}^{-1}X_t$,

$$\begin{aligned}
& \frac{1}{T}\sum_{t=1}^{T}E[\hat{M}_t(u)\hat{M}_t(u)^*] \\
= \quad & \frac{1}{T}\sum_{t=1}^{T}E\left\{\left[X_t e^{iu2\pi t/T} - \hat{Q}_{xx}(u)\hat{Q}_{xx}^{-1}X_t\right]\left[X_t e^{iu2\pi t/T} - \hat{Q}_{xx}(u)^*\hat{Q}_{xx}^{-1}X_t\right]'\right\} \\
= \quad & \frac{1}{T}\sum_{t=1}^{T}E\left\{X_tX_t' - \hat{Q}_{xx}(u)\hat{Q}_{xx}^{-1}X_tX_t'e^{-iu2\pi t/T} - X_tX_t'e^{iu2\pi t/T}\hat{Q}_{xx}^{-1}\hat{Q}_{xx}(u)^* + \hat{Q}_{xx}(u)\hat{Q}_{xx}^{-1}X_tX_t'\hat{Q}_{xx}^{-1}\hat{Q}_{xx}(u)^*\right\} \\
\xrightarrow{p} \quad & Q_{xx} - Q_{xx}(u)Q_{xx}^{-1}Q_{xx}(u)^* \sim \text{p.s.d.}
\end{aligned}$$

Next, we need to show that $s_T(u)$ satisfies the Lipschitz condition. It is equivalent to show that

$$\left\|X_t\varepsilon_t e^{iu_1 2\pi t/T} - X_t\varepsilon_t e^{iu_2 2\pi t/T}\right\| \leq C|u_1 - u_2|,$$

for some $C < \infty$. Observe that by Taylor expansion,

$$\begin{aligned}
& X_t\varepsilon_t e^{iu_1 2\pi t/T} - X_t\varepsilon_t e^{iu_2 2\pi t/T} \\
= \quad & X_t\varepsilon_t\left[e^{i\bar{u}2\pi t/T}2\pi\frac{t}{T}(u_1 - u_2)\right].
\end{aligned}$$

We can choose C to be $2\pi E(\|X_t\|^2)E(\varepsilon_t^2)$ and the Lipschitz condition is satisfied.

Then by Theorem 2 of Hansen (1996), the consistency of the resampling method is established. $\square$

Proof of Theorem 3.2.4. Under $\mathbb{H}_A$, for each fixed $u \in \mathbb{R}$

$$\begin{aligned}\hat{A}(u) &\xrightarrow{p} \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T}\left[I_d e^{iu2\pi t/T} - Q_{xx}(u)Q_{xx}^{-1}(0)\right] E(X_t X_t')\beta_t + Op(T^{-1/2}) \\ &= Q_{xx}(u)[\tilde{\beta}(u) - \tilde{\beta}(0)] + op(1),\end{aligned}$$

where

$$\tilde{\beta}(u) = Q_{xx}^{-1}(u) \lim_{T\to\infty} \sum_{t=1}^{T} E(X_t X_t') e^{iu2\pi t/T}\beta_t.$$

In the proof in Theorem 3.2.1, we established that

$$S_T(u) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} X_t \varepsilon_t e^{iu2\pi t/T}$$

is stochastically equicontinuous over $u \in \mathbb{R}$. Then

$$\sqrt{T}\hat{A}(u) = S_T(u) - \hat{Q}_{xx}(u)\hat{Q}_{xx}^{-1}(0)S_T(0)$$

is also stochastically equicontinuous. Therefore,

$$\begin{aligned}\hat{K} &= T\sup_{u\in\mathbb{R}} \|\hat{A}(u)\|^2 \\ &= T\sup_{u\in\mathbb{R}} \|\hat{Q}_{xx}(u)[\hat{\beta}(u) - \hat{\beta}(0)]\|^2 \\ &= Op(T),\end{aligned}$$

and

$$\begin{aligned}\hat{C} &= T\int_{\mathbb{R}} \|\hat{A}(u)\|^2 W(u)du \\ &= T\int_{\mathbb{R}} \|\hat{Q}_{xx}(u)[\hat{\beta}(u) - \hat{\beta}(0)]\|^2 W(u)du \\ &= Op(T).\end{aligned}$$

Hereby, we have shown $Pr(\hat{K} > c_T) \to 1$ and $Pr(\hat{C} > c_T) \to 1$ as $T \to \infty$ for any sequence of non-stochastic constants $c_T = o(T)$. □

Proof of Theorem 3.2.5. Under $\mathbb{H}_{A1} : \beta_t = \beta_0 + \delta_T \phi_t$,

$$\hat{A}(u) = \frac{1}{T}\sum_{t=1}^{T} X_t \hat{\varepsilon}_t e^{iu2\pi t/T}$$

$$
\begin{aligned}
&= \frac{1}{T}\sum_{t=1}^{T} X_t X_t' e^{iu2\pi t/T}\left[\beta_t - \left(\frac{1}{T}\sum_{t=1}^{T} X_t X_t'\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T} X_t X_t'\beta_t\right)\right] \\
&\quad + \frac{1}{T}\sum_{t=1}^{T}\left[X_t e^{iu2\pi t/T} - \left(\frac{1}{T}\sum_{t=1}^{T} X_t X_t' e^{iu2\pi t/T}\right)\left(\frac{1}{T}\sum_{t=1}^{T} X_t X_t'\right)^{-1} X_t\right]\varepsilon_t \\
&= \frac{1}{T}\sum_{t=1}^{T} X_t X_t' e^{iu2\pi t/T}\left[\beta_0 + \delta_T\phi_t - \left(\frac{1}{T}\sum_{t=1}^{T} X_t X_t'\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T} X_t X_t'[\beta_0 + \delta_T\phi_t]\right)\right] \\
&\quad + \frac{1}{T}\sum_{t=1}^{T}\left[X_t e^{iu2\pi t/T} - \left(\frac{1}{T}\sum_{t=1}^{T} X_t X_t' e^{iu2\pi t/T}\right)\left(\frac{1}{T}\sum_{t=1}^{T} X_t X_t'\right)^{-1} X_t\right]\varepsilon_t \\
&= \frac{\delta_T}{T}\sum_{t=1}^{T} X_t X_t' e^{iu2\pi t/T}\left[\phi_t - \left(\frac{1}{T}\sum_{t=1}^{T} X_t X_t'\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T} X_t X_t'\phi_t\right)\right] \\
&\quad + \frac{1}{T}\sum_{t=1}^{T}\left[X_t e^{iu2\pi t/T} - \left(\frac{1}{T}\sum_{t=1}^{T} X_t X_t' e^{iu2\pi t/T}\right)\left(\frac{1}{T}\sum_{t=1}^{T} X_t X_t'\right)^{-1} X_t\right]\varepsilon_t \\
&\equiv \hat{A}_1(u) + \hat{A}_2(u).
\end{aligned}
$$

Under Theorem 3.2.1, we have $\sqrt{T}\hat{A}_2(u) \Rightarrow G(u)$. $\hat{A}_1(u)$ is the same as $\hat{A}(u)$ under $\mathbb{H}_A$ except that β_t is replaced by $\delta_T\phi_t$. Given $\delta_T = \frac{1}{\sqrt{T}}$,

$$
\begin{aligned}
\sqrt{T}\hat{A}(u) &= \frac{1}{T}\sum_{t=1}^{T} X_t X_t' e^{iu2\pi t/T}\left[\phi_t - \left(\frac{1}{T}\sum_{t=1}^{T} X_t X_t'\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T} X_t X_t'\phi_t\right)\right] \\
&\quad + \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[X_t e^{iu2\pi t/T} - \left(\frac{1}{T}\sum_{t=1}^{T} X_t X_t' e^{iu2\pi t/T}\right)\left(\frac{1}{T}\sum_{t=1}^{T} X_t X_t'\right)^{-1} X_t\right]\varepsilon_t \\
&= \sqrt{T}\hat{A}_1(u) + \sqrt{T}\hat{A}_2(u).
\end{aligned}
$$

By Theorem 3.2.2

$$
\sup_{u\in\mathbb{R}}\left\|\hat{A}_1(u) - \xi(u)\right\| \xrightarrow{p} 0
$$

where

$$
\begin{aligned}
\xi(u) &= \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} E(X_t X_t')\phi_t e^{iu2\pi t/T} \\
&\quad - \lim_{T\to\infty}\left(\frac{1}{T}\sum_{t=1}^{T} E[X_t X_t']e^{iu2\pi t/T}\right)\left(\frac{1}{T}\sum_{t=1}^{T} E[X_t X_t']\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T} E[X_t X_t']\phi_t\right).
\end{aligned}
$$

By Theorem 3.2.1

$$\sqrt{T}\hat{A}_2(u) \Rightarrow G(u),$$

where $G(u)$ is a complex-valued Gaussian process defined in Theorem 3.2.1. Next, we need to show that $\hat{A}_1(u)$ is stochastically equicontinuous. We just need to show that

$$\frac{1}{T}\sum_{t=1}^{T} X_t X_t' \phi_t e^{iu2\pi t/T}$$

is stochastically equicontinuous. Given $E(|X_{jt}|^4) < \infty$ and $\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \|\phi_t^2\| < \infty$, we can prove this result using similar technique as in Theorem 3.2.1. Therefore, under $\mathbb{H}_{A1}$,

$$\sqrt{T}\hat{A}(u) \Rightarrow \xi(u) + G(u),$$

for all $u \neq 0$. By continuous mapping theorem,

$$\begin{aligned} \hat{K} &\xrightarrow{d} \sup_{u\in\mathbb{R}} \|\xi(u) + G(u)\|^2, \\ \hat{C} &\xrightarrow{d} \int_{\mathbb{R}} \|\xi(u) + G(u)\|^2 W(u) du. \end{aligned}$$

□

Proof of Theorem 3.3.1. Under $\mathbb{H}_0 : \beta_t = \beta_0$,

$$\sqrt{T}[\hat{\beta}^{IV}(u) - \beta_0] = \left[\hat{Q}_{xz}\hat{Q}_{zz}^{-1}\hat{Q}_{zx}(u)\right]^{-1} \hat{Q}_{xz}\hat{Q}_{zz}^{-1} S_T^{IV}(u),$$

where

$$S_T^{IV}(u) \equiv \frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_t \varepsilon_t e^{iu2\pi t/T}.$$

To establish the result, we first need to show

- $\sup_{u\in\mathbb{R}} \left\|\hat{Q}_{zx}(u) - Q_{zx}(u)\right\| \xrightarrow{p} 0;$

- $\sup_{u_1,u_2\in\mathbb{R}^2}\left\|\hat{V}_{zz}(u_1,u_2)-V_{zz}(u_1,u_2)\right\|\xrightarrow{p}0.$

Given Assumptions 3.3.1 and 3.3.5, we can prove these using similar arguments as in the proof of Lemma 3.2.1.

Let $B_t\equiv Z_tX_t'-E(Z_tX_t)$ be a $l\times d$ matrix with each entry being $b_{jkt}=Z_{jt}X_{kt}-E(Z_{jt}X_{kt})$, where Z_{jt} and X_{kt} are the jth and kth element respectively in Z_t for $1\leq j\leq l$ and $1\leq k\leq d$. Then

$$\begin{aligned}
& \sup_{u\in\mathbb{R}}\left\|\frac{1}{T}\sum_{t=1}^{T}Z_tX_t'e^{iu2\pi t/T}-\frac{1}{T}\sum_{t=1}^{T}E(Z_tX_t')e^{iu2\pi t/T}\right\| \\
= & \sup_{u\in\mathbb{R}}\left\|\frac{1}{T}\sum_{t=1}^{T}B_te^{iu2\pi t/T}\right\| \\
= & \sup_{u\in\mathbb{R}}\left(\sum_{j=1}^{l}\sum_{k=1}^{d}\left|\frac{1}{T}\sum_{t=1}^{T}b_{jkt}e^{iu2\pi t/T}\right|^2\right)^{1/2} \\
= & \sup_{u\in\mathbb{R}}\left[\sum_{j=1}^{l}\sum_{k=1}^{d}\left(\frac{1}{T^2}\sum_{t=1}^{T}b_{jkt}^2+\frac{1}{T^2}\sum_{s\neq t}^{T}b_{jkt}b_{jks}e^{iu2\pi(t-s)/T}\right)\right]^{1/2} \\
\equiv & R_6,\ say.
\end{aligned}$$

Next, we show that $R_6=op(1)$:

$$\begin{aligned}
ER_1 &= E\left\{\sup_{u\in\mathbb{R}}\left[\sum_{j=1}^{l}\sum_{k=1}^{d}\left(\frac{1}{T^2}\sum_{t=1}^{T}b_{jkt}^2+\frac{1}{T^2}\sum_{s\neq t}^{T}b_{jkt}b_{jks}e^{iu2\pi(t-s)/T}\right)\right]^{1/2}\right\} \\
&\leq \sup_{u\in\mathbb{R}}E\left\{\left[\sum_{j=1}^{l}\sum_{k=1}^{d}\left(\frac{1}{T^2}\sum_{t=1}^{T}b_{jkt}^2+\frac{1}{T^2}\sum_{s\neq t}^{T}b_{jkt}b_{jks}e^{iu2\pi(t-s)/T}\right)\right]^{1/2}\right\} \\
&\leq \sup_{u\in\mathbb{R}}\left\{E\left[\sum_{j=1}^{l}\sum_{k=1}^{d}\left(\frac{1}{T^2}\sum_{t=1}^{T}b_{jkt}^2+\frac{1}{T^2}\sum_{s\neq t}^{T}b_{jkt}b_{jks}e^{iu2\pi(t-s)/T}\right)\right]\right\}^{1/2},
\end{aligned}$$

by Jensen's inequality. It is straightforward to show

$$\begin{aligned}
& \sup_{u\in\mathbb{R}}\left\{E\left[\sum_{j=1}^{l}\sum_{k=1}^{d}\left(\frac{1}{T^2}\sum_{t=1}^{T}b_{jkt}^2+\frac{1}{T^2}\sum_{s\neq t}^{T}b_{jkt}b_{jks}e^{iu2\pi(t-s)/T}\right)\right]\right\}^{1/2} \\
= & \sup_{u\in\mathbb{R}}\left\{\sum_{j=1}^{l}\sum_{k=1}^{d}\left(\frac{1}{T^2}\sum_{t=1}^{T}E(b_{jkt}^2)+\frac{1}{T^2}\sum_{s\neq t}^{T}E(b_{jkt}b_{jks})e^{iu2\pi(t-s)/T}\right)\right\}^{1/2}
\end{aligned}$$

$$
\begin{aligned}
&= \sup_{u\in\mathbb{R}}\left\{\sum_{j=1}^{l}\sum_{k=1}^{d}\left(\frac{1}{T^2}\sum_{t=1}^{T}E(b_{jkt}^2)+\frac{2}{T^2}\sum_{1\le s\le t\le T}^{T}E(b_{jkt}b_{jks})e^{iu2\pi(t-s)/T}\right)\right\}^{1/2}\\
&\le \sup_{u\in\mathbb{R}}\left\{\sum_{j=1}^{l}\sum_{k=1}^{d}\left(\frac{1}{T^2}\sum_{t=1}^{T}E(|Z_{jt}X_{kt}|^2)\right.\right.\\
&\quad \left.\left.+\frac{2}{T^2}\sum_{1\le s\le t\le T}^{T}\alpha(|s-t|)^{\frac{\delta-1}{\delta}}(E|Z_{jt}X_{kt}|^{2\delta})^{\frac{1}{2\delta}}(E|Z_{js}X_{ks}|^{2\delta})^{\frac{1}{2\delta}}e^{iu2\pi(t-s)/T}\right)\right\}^{1/2}\\
&\le \sup_{u\in\mathbb{R}}\left\{\sum_{j=1}^{l}\sum_{k=1}^{d}\left[\frac{1}{T^2}\sum_{t=1}^{T}E(|Z_{jt}X_{kt}|^2+\frac{1}{T}\sum_{m=1}^{T-1}\left(1-\frac{m}{T}\right)\alpha(|m|)^{\frac{\delta-1}{\delta}}(E|Z_{jt}X_{kt}|^{4\delta})^{\frac{1}{\delta}}e^{iu2\pi j/T}\right]\right\}^{1/2}\\
&= O(T^{-1/2}),
\end{aligned}
$$

given $E|X_{jt}|^{4\delta}\le C<\infty$ and $\sum_{j=1}^{\infty}\alpha(j)^{\frac{\delta-1}{\delta}}<C$.

Then, we can obtain

$$
\begin{aligned}
E(R_6^2) &= E\left\{\sup_{u\in\mathbb{R}}\left[\sum_{j,k=1}^{d}\left(\frac{1}{T^2}\sum_{t=1}^{T}b_{jkt}^2+\frac{1}{T^2}\sum_{s\ne t}^{T}b_{jkt}b_{jks}e^{iu2\pi(t-s)/T}\right)\right]\right\}\\
&= \sup_{u\in\mathbb{R}}\left\{E\left[\sum_{j,k=1}^{d}\left(\frac{1}{T^2}\sum_{t=1}^{T}b_{jkt}^2+\frac{1}{T^2}\sum_{s\ne t}^{T}b_{jkt}b_{jks}e^{iu2\pi(t-s)/T}\right)\right]\right\}\\
&= Op(T^{-1}).
\end{aligned}
$$

Thus, by Chebyshev's inequality, $R_6 = op(1)$. Thus, we have shown $\sup_{u\in\mathbb{R}}\left\|\frac{1}{T}\sum_{t=1}^{T}B_te^{iu2\pi t/T}\right\| = op(1)$. By similar derivation, given $E(|\varepsilon_t|^{4\delta})<C<\infty$, we can show

$$
\sup_{u_1,u_2\in\mathbb{R}^2}\left\|\hat{V}_{zz}(u_1,u_2)-V_{zz}(u_1,u_2)\right\|\xrightarrow{p}0,
$$

since we can view $u_1,u_2\in\mathbb{R}^2$ as $v\in\mathbb{R}$ where $v\equiv u_1-u_2$.

Next we show that for any fixed $u\in\mathbb{R}$, $\sqrt{T}[\hat{\beta}^{IV}(u)-\beta_0]$ converges in distribution to a normal distribution under $\mathbb{H}_0$.

For any $u\in\mathbb{R}$,

$$
U_T^{IV}(u) = \tau' S_T^{IV}(u)
$$

$$\begin{aligned} &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \tau' Z_t \varepsilon_t e^{iu2\pi t/T} \\ &= \sum_{t=1}^{T} \left[\frac{1}{\sqrt{T}} \sum_{j=1}^{l} \tau_j Z_{jt} \varepsilon_t e^{iu2\pi t/T} \right], \end{aligned}$$

where $\tau \in \mathbb{R}^l$ is any nonzero vector such that $\tau'\tau = 1$, and τ_j and $Z_{jt}(u)$ represent the jth entry of $l \times 1$ column vectors τ and $Z_t(u)$ respectively. By similar derivation as in Proof of Theorem 3.2.1 and by Theorem 1.1 in Bradley and Tone (2015),

$$S_T^{IV}(u) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} Z_t(u)\varepsilon_t \xrightarrow{d} N[0, V_{zz}(u,u)],$$

for each fixed $u \in \mathbb{R}$, where

$$V_{zz}(u_1, u_2) \quad = \quad \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} E[Z_t(u_1) Z_t(u_2)^{*'} \varepsilon_t^2].$$

Next, we show $S_T^{IV}(u)$ is stochastically equicontinuous. By definition, we need show that for all $\epsilon > 0$ and $\eta > 0$, there exists $\vartheta > 0$ such that

$$\limsup_{T\to\infty} Pr\left(\sup_{u\in\mathbb{R}} \sup_{u'\in B(u,\vartheta)} \|S_T^{IV}(u) - S_T^{IV}(u')\| > \eta \right) < \epsilon,$$

where $B(u, \vartheta)$ is an open set in $\mathbb{R}$ such that $|u' - u| \le \vartheta$ for all $u' \in B(u, \vartheta)$.

By Taylor expansion, there exists an $\bar{u} \in B(u, \vartheta)$ lies in between u and u' such that

$$\begin{aligned} S_T^{IV}(u) &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} Z_t \varepsilon_t e^{iu2\pi t/T} \\ &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} Z_t \varepsilon_t \left[e^{iu'2\pi t/T} + \frac{i2\pi t}{T} e^{i\bar{u}2\pi t/T} (u - u') \right], \end{aligned}$$

then we have

$$\left\| S_T^{IV}(u) - S_T^{IV}(u') \right\| = \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} Z_t \varepsilon_t e^{i\bar{u}2\pi t/T} \frac{i2\pi t}{T} \right\| |u - u'|.$$

Thus

$$
\begin{aligned}
& \limsup_{T\to\infty} Pr\left(\sup_{u\in\mathbb{R}} \sup_{u'\in B(u,\vartheta)} \|S_T^{IV}(u) - S_T^{IV}(u')\| > \eta\right) \\
= \; & \limsup_{T\to\infty} Pr\left(\sup_{u\in\mathbb{R}} \sup_{u'\in B(u,\vartheta)} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} Z_t\varepsilon_t e^{i\bar{u}2\pi t/T} \frac{i2\pi t}{T} \right\| |u-u'| > \eta\right) \\
\leq \; & \frac{1}{\eta^2} \limsup_{T\to\infty} E\left(\sup_{u\in\mathbb{R}} \sup_{u'\in B(u,\vartheta)} \left\| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} Z_t\varepsilon_t e^{i\bar{u}2\pi t/T} \frac{i2\pi t}{T} \right\| |u-u'|\right)^2 \\
= \; & \frac{4\pi^2}{\eta^2} \limsup_{T\to\infty} E\left(\sup_{u\in\mathbb{R}} \sup_{u'\in B(u,\vartheta)} \sqrt{\sum_{j=1}^{l} \frac{1}{T} \sum_{s,t=1}^{T} Z_{jt}Z_{js}\varepsilon_t\varepsilon_s e^{i\bar{u}2\pi(t-s)/T} \frac{st}{T^2} |u-u'|}\right)^2 \\
\leq \; & \frac{4\pi^2}{\eta^2} \limsup_{T\to\infty} E\left(\sup_{u\in\mathbb{R}} \sup_{u'\in B(u,\vartheta)} \sum_{j=1}^{l} \frac{1}{T} \sum_{s,t=1}^{T} Z_{jt}Z_{js}\varepsilon_t\varepsilon_s e^{i\bar{u}2\pi(t-s)/T} \frac{st}{T^2} |u-u'|^2\right) \\
\leq \; & \frac{4\pi^2}{\eta} \vartheta^2 C_2,
\end{aligned}
$$

where

$$
C_2 \geq d \max_j E(\|Z_t\|^4) E(\varepsilon_{jt}^4) < \infty.
$$

It follows that for any $\epsilon > 0$ and $\eta > 0$, there exists $\vartheta < \frac{\sqrt{\eta\epsilon}}{2\pi\sqrt{C_2}}$ such that

$$
\frac{4\pi^2}{\eta} \vartheta^2 C_2 < \epsilon.
$$

Therefore, $S_T^{IV}(u)$ is stochastically equicontinuous over $u \in \mathbb{R}$. Given the uniform convergence for $\hat{Q}_{zx}(u)$ and $\hat{V}_{zz}(u_1, u_2)$,

$$
\sqrt{T}\left[\hat{\beta}^{IV}(u) - \beta_0\right] \Rightarrow G^{IV}(u),
$$

where $G^{IV}(u)$ is a zero-mean complex-valued Gaussian process with covariance kernel

$$
\begin{aligned}
K^{IV}(u_1, u_2) &= cov\left[G^{IV}(u_1), G^{IV}(u_2)^*\right] \\
&= [Q_{xz}Q_{zz}^{-1}Q_{zx}(u_1)]^{-1} Q_{xz}Q_{zz}^{-1} V_{zz}(u_1, u_2) Q_{zz}^{-1} Q_{zx} [Q_{xz}Q_{zz}^{-1}Q_{zx}(u_2)^*]^{-1}.
\end{aligned}
$$

In particular, when $u = 0$,

$$\sqrt{T}\left[\hat{\beta}^{IV}(0) - \beta_0\right] \xrightarrow{d} N[0, K^{IV}(0,0)].$$

Given Assumptions 3.1 to 3.5, under $\mathbb{H}_0$,

$$\begin{aligned}
\sqrt{T}\hat{A}^{IV}(u) &= \hat{Q}_{xz}\hat{Q}_{zz}^{-1}\hat{Q}_{zx}(u)\sqrt{T}\left[\hat{\beta}^{IV}(u) - \hat{\beta}^{IV}(0)\right] \\
&= \hat{Q}_{xz}\hat{Q}_{zz}^{-1}\hat{Q}_{zx}(u)\left\{\sqrt{T}\left[\hat{\beta}^{IV}(u) - \beta_0\right] - \sqrt{T}\left[\hat{\beta}^{IV}(0) - \beta_0\right]\right\} \\
&\Rightarrow Q_{xz}Q_{zz}^{-1}Q_{zx}(u)[G^{IV}(u) - G^{IV}(0)] \\
&= \mathcal{G}^{IV}(u),
\end{aligned}$$

where $\mathcal{G}^{IV}(u)$ is a complex-valued Gaussian process with mean 0 and variance covariance kernel

$$\begin{aligned}
\mathcal{K}^{IV}(u_1,u_2) &= cov[\mathcal{G}^{IV}(u_1), \mathcal{G}^{IV}(u_2)^*] \\
&= cov\left\{Q_{xz}Q_{zz}^{-1}Q_{zx}(u_1)[G(u_1) - G(0)], [G(u_2)^{*'} - G(0)']Q_{xz}(u_2)^*Q_{zz}^{-1}Q_{zx}\right\} \\
&= Q_{xz}Q_{zz}^{-1}Q_{zx}(u_1)cov\left[G(u_1), G(u_2)^{*'}\right]Q_{xz}(u_2)^*Q_{zz}^{-1}Q_{zx} \\
&\quad -Q_{xz}Q_{zz}^{-1}Q_{zx}(u_1)cov\left[G(u_1), G(0)^{*'}\right]Q_{xz}(u_2)^*Q_{zz}^{-1}Q_{zx} \\
&\quad -Q_{xz}Q_{zz}^{-1}Q_{zx}(u_1)cov\left[G(0), G(u_2)^{*'}\right]Q_{xz}(u_2)^*Q_{zz}^{-1}Q_{zx} \\
&\quad +Q_{xz}Q_{zz}^{-1}Q_{zx}(u_1)cov\left[G(0), G(0)^{*'}\right]Q_{xz}(u_2)^*Q_{zz}^{-1}Q_{zx} \\
&= Q_{xz}Q_{zz}^{-1}Q_{zx}(u_1)\left[K(u_1,u_2) - K(0,u_2) - K(u_1,0) + K(0,0)\right]Q_{xz}(u_2)^*Q_{zz}^{-1}Q_{zx}.
\end{aligned}$$

□

Proof of Theorem 3.3.2. Under $\mathbb{H}_A$,

$$\begin{aligned}
\hat{\beta}^{IV}(u) &= [\hat{Q}_{xz}\hat{Q}_{zz}^{-1}\hat{Q}_{zx}(u)]^{-1}\hat{Q}_{xz}\hat{Q}_{zz}^{-1}\frac{1}{T}\sum_{t=1}^{T} Z_t(u)Y_t \\
&= [\hat{Q}_{xz}\hat{Q}_{zz}^{-1}\hat{Q}_{zx}(u)]^{-1}\hat{Q}_{xz}\hat{Q}_{zz}^{-1}\frac{1}{T}\sum_{t=1}^{T} Z_t(u)X_t'\beta_t
\end{aligned}$$

$$
\begin{aligned}
&+[\hat{Q}_{xz}\hat{Q}_{zz}^{-1}\hat{Q}_{zx}(u)]^{-1}\hat{Q}_{xz}\hat{Q}_{zz}^{-1}\frac{1}{T}\sum_{t=1}^{T}Z_t(u)\varepsilon_t \\
= \quad &[\hat{Q}_{xz}\hat{Q}_{zz}^{-1}\hat{Q}_{zx}(u)]^{-1}\hat{Q}_{xz}\hat{Q}_{zz}^{-1}\frac{1}{T}\sum_{t=1}^{T}Z_t(u)X_t'\beta_t + op(1),
\end{aligned}
$$

where the last equality comes form the result in Theorem 3.3.1. We first show

$$
\sup_{u\in\mathbb{R}}\left\|\frac{1}{T}\sum_{t=1}^{T}Z_tX_t'\beta_t e^{iu2\pi t/T} - \frac{1}{T}\sum_{t=1}^{T}E(Z_tX_t')\beta_t e^{iu2\pi t/T}\right\| \xrightarrow{p} 0,
$$

as $T\to\infty$. Following the same notation as in Proof of Theorem 3.3.1, we have

$$
\begin{aligned}
&\sup_{u\in\mathbb{R}}\left\|\frac{1}{T}\sum_{t=1}^{T}Z_tX_t'\beta_t e^{iu2\pi t/T} - \frac{1}{T}\sum_{t=1}^{T}E(Z_tX_t')\beta_t e^{iu2\pi t/T}\right\| \\
= \quad &\sup_{u\in\mathbb{R}}\left\|\frac{1}{T}\sum_{t=1}^{T}B_t\beta_t e^{iu2\pi t/T}\right\| \\
= \quad &\sup_{u\in\mathbb{R}}\left[\sum_{j=1}^{l}\left|\frac{1}{T}\sum_{t=1}^{T}\sum_{k=1}^{d}b_{jkt}\beta_{jt}e^{iu2\pi t/T}\right|^2\right]^{1/2} \\
= \quad &\sup_{u\in\mathbb{R}}\left\{\sum_{j=1}^{l}\left[\frac{1}{T^2}\sum_{t=1}^{T}\left(\sum_{k=1}^{d}b_{jkt}\beta_{jt}\right)^2 + \frac{1}{T^2}\sum_{s\neq t}^{T}\left(\sum_{k_1,k_2=1}^{d}b_{jk_1t}b_{jk_2s}\beta_{jt}\beta_{js}\right)e^{iu2\pi(t-s)/T}\right]\right\}^{1/2} \\
\equiv \quad &R_7,\ say.
\end{aligned}
$$

Then

$$
\begin{aligned}
ER_7 \quad \leq \quad &\sup_{u\in\mathbb{R}}E\left\{\sum_{j=1}^{l}\left[\frac{1}{T^2}\sum_{t=1}^{T}\left(\sum_{k=1}^{d}b_{jkt}\beta_{jt}\right)^2 + \frac{1}{T^2}\sum_{s\neq t}^{T}\left(\sum_{k_1,k_2=1}^{d}b_{jk_1t}b_{jk_2s}\beta_{jt}\beta_{js}\right)e^{iu2\pi(t-s)/T}\right]\right\}^{1/2} \\
\leq \quad &\sup_{u\in\mathbb{R}}\left\{\sum_{j=1}^{l}E\left[\frac{1}{T^2}\sum_{t=1}^{T}\left(\sum_{k=1}^{d}b_{jkt}\beta_{jt}\right)^2 + \frac{1}{T^2}\sum_{s\neq t}^{T}\left(\sum_{k_1,k_2=1}^{d}b_{jk_1t}b_{jk_2s}\beta_{jt}\beta_{js}\right)e^{iu2\pi(t-s)/T}\right]\right\}^{1/2} \\
= \quad &\sup_{u\in\mathbb{R}}\left\{\sum_{j=1}^{l}\left[\frac{1}{T^2}\sum_{t=1}^{T}E\left(\sum_{k=1}^{d}b_{jkt}\beta_{jt}\right)^2 + \frac{1}{T^2}\sum_{s\neq t}^{T}E\left(\sum_{k_1,k_2=1}^{d}b_{jk_1t}b_{jk_2s}\beta_{jt}\beta_{js}\right)e^{iu2\pi(t-s)/T}\right]\right\}^{1/2} \\
= \quad &Op(T^{-1/2}),
\end{aligned}
$$

given similar argument as proof in R_6 and the fact that $\sum_{t=1}^{T}\|\beta_t\|^2 < C < \infty$ We show that $R_7 = op(1)$. Similarly, we can show $ER_7^2 = Op(T^{-1})$ and thus $R_2 = op(1)$ by Chebyshev's inequality.

Also, given $\sup_{u\in\mathbb{R}} \left\|\hat{Q}_{zx}(u) - Q_{zx}(u)\right\| \xrightarrow{p} 0$, thus we have

$$sup_{u\in\mathbb{R}} \left\|\hat{\beta}^{IV}(u) - \bar{\beta}^{IV}(u)\right\| \xrightarrow{p} 0,$$

where $\tilde{\beta}^{IV}(u) = [Q_{xz}Q_{zz}^{-1}Q_{zx}(u)]^{-1}Q_{xz}Q_{zz}^{-1}\lim_{T\to}\frac{1}{T}\sum_{t=1}^{T}E(Z_tX_t')e^{iu2\pi t/T}\beta_t$.

Given

$$\begin{aligned}
\hat{A}^{IV}(u) &= \hat{Q}_{xz}\hat{Q}_{zz}^{-1}\hat{Q}_{zx}(u)[\hat{\beta}^{IV}(u) - \hat{\beta}^{IV}(0)] \\
&= \hat{Q}_{xz}\hat{Q}_{zz}^{-1}\frac{1}{T}\sum_{t=1}^{T}Z_t(u)Y_t - \hat{Q}_{xz}\hat{Q}_{zz}^{-1}\hat{Q}_{zx}(u)\hat{\beta}^{IV}(0) \\
&= \hat{Q}_{xz}\hat{Q}_{zz}^{-1}\frac{1}{T}\sum_{t=1}^{T}Z_t(u)X_t'\beta_t - \hat{Q}_{xz}\hat{Q}_{zz}^{-1}\hat{Q}_{zx}(u)\hat{\beta}^{IV}(0) + op(1) \\
&= \hat{Q}_{xz}\hat{Q}_{zz}^{-1}\frac{1}{T}\sum_{t=1}^{T}[Z_tX_t' - E(Z_tX_t')]\beta_te^{iu2\pi t/T} + \hat{Q}_{xz}\hat{Q}_{zz}^{-1}\frac{1}{T}\sum_{t=1}^{T}E(Z_tX_t')\beta_te^{iu2\pi t/T} \\
&\quad - \left\{[\hat{Q}_{xz}\hat{Q}_{zz}^{-1}\hat{Q}_{zx}(u) - Q_{xz}Q_{zz}^{-1}Q_{zx}(u)]\hat{\beta}^{IV}(0) + Q_{xz}Q_{zz}^{-1}Q_{zx}(u)\hat{\beta}^{IV}(0)\right\} + op(1),
\end{aligned}$$

then

$$\begin{aligned}
&\sup_{u\in\mathbb{R}}\left\|\hat{A}^{IV}(u) - \frac{1}{T}\sum_{t=1}^{T}\left[I_de^{iu2\pi t/T} - Q_{xz}Q_{zz}^{-1}Q_{zx}(u)[Q_{xz}Q_{zz}^{-1}Q_{zx}(0)]^{-1}\right]Q_{xz}Q_{zz}^{-1}E(Z_tX_t')\beta_t\right\| \\
=\ &\sup_{u\in\mathbb{R}}\left\|\hat{A}^{IV}(u) - Q_{xz}Q_{zz}^{-1}\frac{1}{T}\sum_{t=1}^{T}E(Z_tX_t')\beta_te^{iu2\pi t/T}\right. \\
&\left.+Q_{xz}Q_{zz}^{-1}Q_{zx}(u)[Q_{xz}Q_{zz}^{-1}Q_{zx}(0)]^{-1}Q_{xz}Q_{zz}^{-1}\sum_{t=1}^{T}E(Z_tX_t')\beta_t\right\| \\
=\ &\sup_{u\in\mathbb{R}}\left\|\hat{Q}_{xz}\hat{Q}_{zz}^{-1}\frac{1}{T}\sum_{t=1}^{T}[Z_tX_t' - E(Z_tX_t')]\beta_te^{iu2\pi t/T} + [Q_{xz}Q_{zz}^{-1} - \hat{Q}_{xz}\hat{Q}_{zz}^{-1}]\sum_{t=1}^{T}E(Z_tX_t')\beta_te^{iu2\pi t/T}\right. \\
&-[\hat{Q}_{xz}\hat{Q}_{zz}^{-1}\hat{Q}_{zx}(u) - Q_{xz}Q_{zz}^{-1}Q_{zx}(u)]\hat{\beta}^{IV}(0) \\
&\left.+[Q_{xz}Q_{zz}^{-1}Q_{zx}(u)][Q_{xz}Q_{zz}^{-1}Q_{zx}(0)]^{-1}Q_{xz}Q_{zz}^{-1}\sum_{t=1}^{T}E(Z_tX_t')\beta_t - \hat{\beta}^{IV}(0)]]\right\| \\
\leq\ &\hat{Q}_{xz}\hat{Q}_{zz}^{-1}\sup_{u\in\mathbb{R}}\left\|\frac{1}{T}\sum_{t=1}^{T}[Z_tX_t' - E(Z_tX_t')]\beta_te^{iu2\pi t/T}\right\| \\
&+[Q_{xz}Q_{zz}^{-1} - \hat{Q}_{xz}\hat{Q}_{zz}^{-1}]\sup_{u\in\mathbb{R}}\left\|\frac{1}{T}\sum_{t=1}^{T}E(Z_tX_t')\beta_te^{iu2\pi t/T}\right\|
\end{aligned}$$

$$
\begin{aligned}
&+\sup_{u\in\mathbb{R}}\left\|[\hat{Q}_{xz}\hat{Q}_{zz}^{-1}\hat{Q}_{zx}(u)-Q_{xz}Q_{zz}^{-1}Q_{zx}(u)]\right\|\hat{\beta}^{IV}(0)\\
&+\sup_{u\in\mathbb{R}}\left\|[Q_{xz}Q_{zz}^{-1}Q_{zx}(u)][Q_{xz}Q_{zz}^{-1}Q_{zx}(0)]^{-1}Q_{xz}Q_{zz}^{-1}\sum_{t=1}^{T}E(Z_tX_t')\beta_t-\hat{\beta}^{IV}(0)]]\right\|\\
=\;\; & R_8+R_9+R_{10}+R_{11},
\end{aligned}
$$

by triangle inequality. First, $R_8=op(1)$ by previous results. $R_9=op(1)$ because $\hat{Q}_{xz}\xrightarrow{p}Q_{xz}$ and $\hat{Q}_{zz}\xrightarrow{p}Q_{zz}$ by the regularity conditions and Assumption 3.6. By the same result as in the proof of Theorem 3.3.1, we have $R_{10}=op(1)$. $R_{11}=op(1)$ follows from the standard result in the 2SLS estimator. Thus, we have proved the Theorem. □

Proof of Theorem 3.3.3. Theorem 3.3.1 established the weak convergence of $\sqrt{T}\hat{A}^{IV}(u)$ to a complex-valued Gaussian process $\mathcal{G}^{IV}(u)$. By the continuous mapping theorem, we have

$$
\hat{K}^{IV}\xrightarrow{d}\sup_{u\in\mathbb{R}}\|\mathcal{G}^{IV}(u)\|^2,
$$

and

$$
\hat{C}^{IV}\xrightarrow{d}\int_{\mathbb{R}}\|\mathcal{G}^{IV}(u)\|^2W(u)du.
$$

□

Proof of Theorem 3.3.4. Under $\mathbb{H}_A$, for each fixed $u\in\mathbb{R}$

$$
\begin{aligned}
\hat{A}^{IV}(u)\;\;&\xrightarrow{p}\;\;\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\left[I_de^{iu2\pi t/T}-Q_{xz}Q_{zz}^{-1}Q_{zx}(u)[Q_{xz}Q_{zz}^{-1}Q_{zx}(0)]^{-1}\right]Q_{xz}Q_{zz}^{-1}E(Z_tX_t')\beta_t+Op(T^{-1/2})\\
&=\;\;Q_{xz}Q_{zz}^{-1}Q_{zx}(u)[\tilde{\beta}^{IV}(u)-\tilde{\beta}^{IV}(0)]+op(1),
\end{aligned}
$$

where

$$
\tilde{\beta}^{IV}(u)=[Q_{xz}Q_{zz}^{-1}Q_{zx}(u)]^{-1}Q_{xz}Q_{zz}^{-1}\lim_{T\to}\frac{1}{T}\sum_{t=1}^{T}E(Z_tX_t')e^{iu2\pi t/T}\beta_t.
$$

In the proof in Theorem 3.3.1, we established that

$$S_T^{IV}(u) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} Z_t\varepsilon_t e^{iu2\pi t/T}$$

is stochastically equicontinuous over $u \in \mathbb{R}$, then

$$\sqrt{T}\hat{A}^{IV}(u) = S_T^{IV}(u) - \hat{Q}_{xz}\hat{Q}_{zz}^{-1}\hat{Q}_{zx}(u)[\hat{Q}_{xz}\hat{Q}_{zz}^{-1}\hat{Q}_{zx}(0)]^{-1}S_T^{IV}(0)$$

is also stochastically equicontinuous. Therefore,

$$\begin{aligned}
\hat{K}^{IV} &= \sqrt{T}\sup_{u\in\mathbb{R}} \|\hat{A}^{IV}(u)\|^2 \\
&= \sqrt{T}\sup_{u\in\mathbb{R}} \|\hat{Q}_{xz}\hat{Q}_{zz}^{-1}\hat{Q}_{zx}(u)[\hat{\beta}^{IV}(u) - \hat{\beta}^{IV}(0)]\|^2 \\
&= Op(T),
\end{aligned}$$

and

$$\begin{aligned}
\hat{C}^{IV} &= T\int_{\mathbb{R}} \|\hat{A}^{IV}(u)\|^2 W(u)du \\
&= T\int_{\mathbb{R}} \|\hat{Q}_{xz}\hat{Q}_{zz}^{-1}\hat{Q}_{zx}(u)[\hat{\beta}^{IV}(u) - \hat{\beta}^{IV}(0)]\|^2 W(u)du \\
&= Op(T).
\end{aligned}$$

Thus, we have shown $Pr(\hat{K}^{IV} > c_T) \to 1$ and $Pr(\hat{C}^{IV} > c_T) \to 1$ as $T \to \infty$ for any sequence of non-stochastic constants $c_T = o(T)$. □

Proof of Theorem 3.3.5. Under $\mathbb{H}_{A1} : \beta_t = \beta_0 + \delta_T\phi_t$,

$$\begin{aligned}
\hat{A}^{IV}(u) &= \frac{1}{T}\sum_{t=1}^{T} \hat{X}_t\hat{\varepsilon}_t e^{iu2\pi t/T} \\
&= \frac{1}{T}\sum_{t=1}^{T} \hat{X}_t X_t' e^{iu2\pi t/T}\left[\beta_t - \left(\frac{1}{T}\sum_{t=1}^{T}\hat{X}_t\hat{X}_t'\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\hat{X}_t X_t'\beta_t\right)\right] \\
&\quad + \frac{1}{T}\sum_{t=1}^{T}\left[\hat{X}_t e^{iu2\pi t/T} - \left(\frac{1}{T}\sum_{t=1}^{T}\hat{X}_t X_t' e^{iu2\pi t/T}\right)\left(\frac{1}{T}\sum_{t=1}^{T}\hat{X}_t\hat{X}_t'\right)^{-1}\hat{X}_t\right]\varepsilon_t
\end{aligned}$$

$$
\begin{aligned}
&= \hat{Q}_{xz}\hat{Q}_{zz}^{-1}\frac{1}{T}\sum_{t=1}^{T} Z_t X_t' e^{iu2\pi t/T}\left[\beta_t - \left(\hat{Q}_{xz}\hat{Q}_{zz}^{-1}\hat{Q}_{zx}\right)^{-1}\hat{Q}_{xz}\hat{Q}_{zz}^{-1}\left(\frac{1}{T}\sum_{t=1}^{T} Z_t X_t'\beta_t\right)\right] \\
&\quad + \frac{1}{T}\sum_{t=1}^{T}\left[\hat{Q}_{xz}\hat{Q}_{zz}^{-1}Z_t e^{iu2\pi t/T} - \hat{Q}_{xz}\hat{Q}_{zz}^{-1}\hat{Q}_{zx}(u)\left[\hat{Q}_{xz}\hat{Q}_{zz}^{-1}\hat{Q}_{zx}(0)\right]^{-1}\hat{Q}_{xz}\hat{Q}_{zz}^{-1}Z_t\right]\varepsilon_t \\
&= \hat{Q}_{xz}\hat{Q}_{zz}^{-1}\frac{1}{T}\sum_{t=1}^{T} Z_t X_t' e^{iu2\pi t/T}\left[\beta_0 + \delta_T\phi_t - \left(\hat{Q}_{xz}\hat{Q}_{zz}^{-1}\hat{Q}_{zx}\right)^{-1}\hat{Q}_{xz}\hat{Q}_{zz}^{-1}\left(\frac{1}{T}\sum_{t=1}^{T} Z_t X_t'[\beta_0 + \delta_T\phi_t]\right)\right] \\
&\quad + \frac{1}{T}\sum_{t=1}^{T}\left[\hat{Q}_{xz}\hat{Q}_{zz}^{-1}Z_t e^{iu2\pi t/T} - \hat{Q}_{xz}\hat{Q}_{zz}^{-1}\hat{Q}_{zx}(u)\left[\hat{Q}_{xz}\hat{Q}_{zz}^{-1}\hat{Q}_{zx}(0)\right]^{-1}\hat{Q}_{xz}\hat{Q}_{zz}^{-1}Z_t\right]\varepsilon_t \\
&= \hat{Q}_{xz}\hat{Q}_{zz}^{-1}\frac{\delta_T}{T}\sum_{t=1}^{T} Z_t X_t' e^{iu2\pi t/T}\left[\phi_t - \left(\hat{Q}_{xz}\hat{Q}_{zz}^{-1}\hat{Q}_{zx}\right)^{-1}\hat{Q}_{xz}\hat{Q}_{zz}^{-1}\left(\frac{1}{T}\sum_{t=1}^{T} Z_t X_t'\phi_t\right)\right] \\
&\quad + \frac{1}{T}\sum_{t=1}^{T}\left[\hat{Q}_{xz}\hat{Q}_{zz}^{-1}Z_t e^{iu2\pi t/T} - \hat{Q}_{xz}\hat{Q}_{zz}^{-1}\hat{Q}_{zx}(u)\left[\hat{Q}_{xz}\hat{Q}_{zz}^{-1}\hat{Q}_{zx}(0)\right]^{-1}\hat{Q}_{xz}\hat{Q}_{zz}^{-1}Z_t\right]\varepsilon_t \\
&\equiv \hat{A}_1^{IV}(u) + \hat{A}_2^{IV}(u).
\end{aligned}
$$

Under Theorem 3.3.1, we have $\sqrt{T}\hat{A}_2^{IV}(u) \Rightarrow \mathcal{G}^{IV}(u)$. $\hat{A}_1^{IV}(u)$ contains the same as $\hat{A}^{IV}(u)$ under $\mathbb{H}_A$ except that β_t is replaced by $\delta_T\phi_t$. Given $\delta_T = \frac{1}{\sqrt{T}}$,

$$
\begin{aligned}
\sqrt{T}\hat{A}^{IV}(u) &= \hat{Q}_{xz}\hat{Q}_{zz}^{-1}\frac{1}{T}\sum_{t=1}^{T} Z_t X_t' e^{iu2\pi t/T}\left[\phi_t - \left(\hat{Q}_{xz}\hat{Q}_{zz}^{-1}\hat{Q}_{zx}\right)^{-1}\hat{Q}_{xz}\hat{Q}_{zz}^{-1}\left(\frac{1}{T}\sum_{t=1}^{T} Z_t X_t'\phi_t\right)\right] \\
&\quad + \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\hat{Q}_{xz}\hat{Q}_{zz}^{-1}Z_t e^{iu2\pi t/T} - \hat{Q}_{xz}\hat{Q}_{zz}^{-1}\hat{Q}_{zx}(u)\left[\hat{Q}_{xz}\hat{Q}_{zz}^{-1}\hat{Q}_{zx}(0)\right]^{-1}\hat{Q}_{xz}\hat{Q}_{zz}^{-1}Z_t\right]\varepsilon_t \\
&= \sqrt{T}\hat{A}_1(u) + \sqrt{T}\hat{A}_2(u).
\end{aligned}
$$

By Theorem 3.3.2

$$\sup_{u\in\mathbb{R}}\left\|\hat{A}_1^{IV}(u) - \xi^{IV}(u)\right\| \xrightarrow{p} 0$$

where

$$\xi^{IV}(u) = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\left[I_d e^{iu2\pi t/T} - Q_{xz}Q_{zz}^{-1}Q_{zx}(u)[Q_{xz}Q_{zz}^{-1}Q_{zx}(0)]^{-1}\right]Q_{xz}Q_{zz}^{-1}E(Z_tX_t')\phi_t.$$

By Theorem 3.3.1

$$\sqrt{T}\hat{A}_2^{IV}(u) \Rightarrow \mathcal{G}^{IV}(u),$$

where $\mathcal{G}^{IV}(u)$ is a complex-valued Gaussian process defined in Theorem 3.3.1. Next, we need to show that $\hat{A}_1^{IV}(u)$ is stochastically equicontinuous. We just need to show that

$$\frac{1}{T}\sum_{t=1}^{T} Z_t X_t' \phi_t e^{iu2\pi t/T}$$

is stochastically equicontinuous. Given $E(|X_{kt}|^2) < \infty$, $E(|Z_{jt}|^2) < \infty$ for all $j = 1, 2, .., l$ and $k = 1, 2..., d$, and $\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \|\phi_t^2\| < \infty$, we can prove this result using similar procedures as in Theorem 3.3.1. Therefore, under $\mathbb{H}_{A1}$,

$$\sqrt{T}\hat{A}^{IV}(u) \Rightarrow \xi^{IV}(u) + \mathcal{G}^{IV}(u),$$

for all $u \neq 0$. By the continuous mapping theorem,

$$\begin{aligned}
\hat{K}^{IV} &\xrightarrow{d} \sup_{u\in\mathbb{R}} \|\xi^{IV}(u) + \mathcal{G}^{IV}(u)\|^2, \\
\hat{C}^{IV} &\xrightarrow{d} \int_{\mathbb{R}} \|\xi^{IV}(u) + \mathcal{G}^{IV}(u)\|^2 W(u) du.
\end{aligned}$$

□

APPENDIX C

APPENDIX OF CHAPTER 4

C.1 Figures and Tables

Table C.1: Empirical size

		1%	5%	10%
DGP S.1	$n = 100$	0.014	0.065	0.088
	$n = 200$	0.021	0.061	0.092
	$n = 500$	0.012	0.053	0.099
DGP S.2	$n = 100$	0.035	0.093	0.145
	$n = 200$	0.024	0.068	0.122
	$n = 500$	0.021	0.059	0.112
DGP S.3	$n = 100$	0.015	0.065	0.120
	$n = 200$	0.027	0.063	0.112
	$n = 500$	0.025	0.058	0.104

Notes: (i). n is the sample size; (ii). Number of replications = 1000; (iii). $a(X_t) = 1$ for all X_t; (iv). Kernel function is the standard normal density; (v). Bandwidth $h = cn^{-4/17}$ and $c = std(X_t)$; (vi). The critical values are based on the $N(0, 1)$ distribution.

Table C.2: Empirical power

		1%	5%	10%
	$n = 100$	0.822	0.996	0.999
DGP P.1	$n = 200$	0.992	1.000	1.000
	$n = 500$	0.989	1.000	1.000
	$n = 100$	0.884	0.923	0.989
DGP P.2	$n = 200$	0.928	0.978	1.000
	$n = 500$	0.947	1.000	1.000
	$n = 100$	0.929	0.934	0.977
DGP P.3	$n = 200$	0.945	1.000	1.000
	$n = 500$	1.000	1.000	1.000

Notes: (i). n is the sample size; (ii). Number of replications = 1000; (iii). $a(X_t) = 1$ for all X_t; (iv). Kernel function is the standard normal density; (v). Bandwidth $h = cn^{-4/17}$ and $c = std(X_t)$; (vi). The critical values are based on the $N(0, 1)$ distribution.

C.2 Mathematical Proofs

Throughout the appendix, we let C be a generic bounded constant and denote A^* to be the complex conjugate of A, and $Re(A)$ be the real part of A. In addition, we define $P_{sr} \equiv |Y_s - Y_r|_p$ and $Q_{sr} \equiv |Z_s - Z_r|_q$, where $|\cdot|_s$ denotes the Euclidean norm in $\mathbb{R}^s$ for $s = p, q$. Let $\xi_t = (X_t', Y_t', Z_t')'$, define

$$\begin{aligned} \varepsilon_{yz}(u, v, X_s) &= e^{iu'Y_s + iv'Z_s} - \phi_{yz}(u, v, X_s), \\ \varepsilon_y(u, X_s) &= e^{iuY_s} - \phi_y(u, X_s), \\ \varepsilon_z(v, X_s) &= e^{ivZ_s} - \phi_z(v, X_s), \end{aligned}$$

and

$$
\begin{aligned}
\bar{\phi}_{yz}(u,v,x) &= E\hat{\phi}_{yz}(u,v,x), \\
\bar{\phi}_{yz}(u,x) &= E\hat{\phi}_{y}(u,x), \\
\bar{\phi}_{z}(v,x) &= E\hat{\phi}_{z}(v,x).
\end{aligned}
$$

Proof of Theorem 4.3.1. Under $\mathbb{H}_0 : \sigma(u,v,X_t) = \sigma(u,v)$, we can decompose $nh^{k/2}\hat{T}_n$ as

$$
\begin{aligned}
nh^{k/2}\hat{T}_n &= h^{k/2}\sum_{t=1}^{n}\iint \left|\hat{\sigma}(u,v,X_t) - \sigma(u,v,X_t) - [\hat{\sigma}(u,v) - \sigma(u,v)]\right|^2 a(X_t)dW(u,v) \\
&= h^{k/2}\sum_{t=1}^{n}\iint |\hat{\sigma}(u,v,X_t) - \sigma(u,v,X_t)|^2 a(X_t)dW(u,v) \\
&\quad + h^{k/2}\sum_{t=1}^{n}\iint |\hat{\sigma}(u,v) - \sigma(u,v)|^2 a(X_t)dW(u,v) \\
&\quad - 2h^{k/2}\sum_{t=1}^{n}\iint Re\left\{[\hat{\sigma}(u,v,X_t) - \sigma(u,v,X_t)][\hat{\sigma}(u,v) - \sigma(u,v)]^*\right\} a(X_t)dW(u,v) \\
&= T_1 + T_2 - 2T_3, \text{say.}
\end{aligned}
$$

First, let's work on T_1 :

$$
\begin{aligned}
T_1 &= h^{k/2}\sum_{t=1}^{n}\iint |\hat{\sigma}(u,v,X_t) - \sigma(u,v,X_t)|^2 a(X_t)dW(u,v) \\
&= h^{k/2}\sum_{t=1}^{n}\iint \left|\hat{\phi}_{yz}(u,v,X_t) - \hat{\phi}_{y}(u,X_t)\hat{\phi}_{z}(v,X_t) - \sigma(u,v,X_t)\right|^2 a(X_t)dW(u,v) \\
&= h^{k/2}\sum_{t=1}^{n}\iint \Big|\hat{\phi}_{yz}(u,v,X_t) - \bar{\phi}_{yz}(u,v,X_t) - \bar{\phi}_{y}(u,X_t)\left[\hat{\phi}_{z}(v,X_t) - \bar{\phi}_{z}(v,X_t)\right] \\
&\quad - \bar{\phi}_{z}(v,X_t)\left[\hat{\phi}_{y}(u,X_t) - \bar{\phi}_{y}(u,X_t)\right] - \left[\hat{\phi}_{y}(u,X_t) - \bar{\phi}_{y}(u,X_t)\right]\left[\hat{\phi}_{z}(v,X_t) - \bar{\phi}_{z}(v,X_t)\right] \\
&\quad + \bar{\sigma}(u,v,X_t) - \sigma(u,v,X_t)\Big|^2 a(X_t)dW(u,v) \\
&= h^{k/2}\sum_{t=1}^{n}\iint \Big|\hat{\phi}_{yz}(u,v,X_t) - \bar{\phi}_{yz}(u,v,X_t) - \bar{\phi}_{y}(u,X_t)\left[\hat{\phi}_{z}(v,X_t) - \bar{\phi}_{z}(v,X_t)\right] \\
&\quad - \bar{\phi}_{z}(v,X_t)\left[\hat{\phi}_{y}(u,X_t) - \bar{\phi}_{y}(u,X_t)\right] - R_1 + R_2\Big|^2 a(X_t)dW(u,v),
\end{aligned}
$$

where $R_1 = \left[\hat{\phi}_y(u, X_t) - \bar{\phi}_y(u, X_t)\right]\left[\hat{\phi}_z(v, X_t) - \bar{\phi}_z(v, X_t)\right]$, and $R_2 = \bar{\sigma}(u, v, X_t) - \sigma(u, v, X_t)$. Therefore, we have

$$\begin{aligned} T_1 &= h^{k/2}\sum_{t=1}^{n}\iint\Big|\hat{\phi}_{yz}(u, v, X_t) - \bar{\phi}_{yz}(u, v, X_t) - \bar{\phi}_y(u, X_t)\left[\hat{\phi}_z(v, X_t) - \bar{\phi}_z(v, X_t)\right] \\ &\quad -\bar{\phi}_z(v, X_t)\left[\hat{\phi}_y(u, X_t) - \bar{\phi}_y(u, X_t)\right]\Big|^2 a(X_t)dW(u, v) + op(1). \end{aligned}$$

Furthermore, by definition of $\hat{\phi}_{yz}(u, v, X_t)$ and $\bar{\phi}_{yz}(u, v, X_t)$,

$$\begin{aligned} T_1 &= h^{k/2}\sum_{t=1}^{n}\iint\Big|\sum_{s=1}^{n}\varepsilon_{yz}(u, v, X_s)\hat{H}_{st} - \bar{\phi}_y(u, X_t)\sum_{s=1}^{n}\varepsilon_z(v, X_s)\hat{H}_{st} \\ &\quad -\bar{\phi}_z(v, X_t)\sum_{s=1}^{n}\varepsilon_y(u, X_s)\hat{H}_{st}\Big|^2 a(X_t)dW(u, v) + op(1) \\ &= h^{k/2}\sum_{t=1}^{n}\iint\Big|\sum_{s=1}^{n}\hat{H}_{st}\left[\varepsilon_{yz}(u, v, X_s) - \bar{\phi}_y(u, X_t)\varepsilon_z(v, X_s) - \bar{\phi}_z(v, X_t)\varepsilon_y(u, X_s)\right]\Big|^2 a(X_t)dW(u, v) \\ &\quad +op(1) \\ &= h^{k/2}\sum_{t=1}^{n}\iint\Big|\sum_{s=1}^{n}\frac{K(\frac{X_s-X_t}{h})}{nh^k f(X_t)}\left[\varepsilon_{yz}(u, v, X_t) - \bar{\phi}_y(u, X_t)\varepsilon_z(v, X_t) - \bar{\phi}_z(v, X_t)\varepsilon_y(u, X_t)\right]\Big|^2 a(X_t)dW(u, v) \\ &\quad +op(1), \end{aligned}$$

where we use the result of effective kernel provided in Eq.(4.10). We can then further decompose T_1 as

$$\begin{aligned} T_1 &= \frac{1}{n^2h^{3k/2}}\sum_{t=1}^{n}\frac{a(X_t)}{f(X_t)^2}K^2(0)\iint\left|\varepsilon_{yz}(u, v, X_t) - \bar{\phi}_y(u, X_t)\varepsilon_z(v, X_t) - \bar{\phi}_z(v, X_t)\varepsilon_y(u, X_t)\right|^2 dW(u, v) \\ &\quad +\frac{1}{n^2h^{3k/2}}\sum_{s\neq t}^{n}\frac{a(X_t)}{f(X_t)^2}\iint\left|\varepsilon_{yz}(u, v, X_s) - \bar{\phi}_y(u, X_t)\varepsilon_z(v, X_s) - \bar{\phi}_z(v, X_t)\varepsilon_y(u, X_s)\right|^2 dW(u, v) \\ &\quad \times K^2\left(\frac{X_s - X_t}{h}\right) \\ &\quad +\frac{1}{n^2h^{3k/2}}\sum_{s\neq t}^{n}\frac{a(X_t)}{f(X_t)^2}\iint Re\Big\{\left[\varepsilon_{yz}(u, v, X_s) - \bar{\phi}_y(u, X_t)\varepsilon_z(v, X_s) - \bar{\phi}_z(v, X_t)\varepsilon_y(u, X_s)\right] \\ &\quad \times\left[\varepsilon_{yz}(u, v, X_t) - \bar{\phi}_y(u, X_t)\varepsilon_z(v, X_t) - \bar{\phi}_z(v, X_t)\varepsilon_y(u, X_t)\right]^*\Big\}dW(u, v)K(0)K\left(\frac{X_s - X_t}{h}\right) \\ &\quad +\frac{1}{n^2h^{3k/2}}\sum_{s\neq t\neq r}^{n}\frac{a(X_t)}{f(X_t)^2}\iint Re\Big\{\left[\varepsilon_{yz}(u, v, X_s) - \bar{\phi}_y(u, X_t)\varepsilon_z(v, X_s) - \bar{\phi}_z(v, X_t)\varepsilon_y(u, X_s)\right] \end{aligned}$$

$$
\begin{aligned}
&\times\Big[\varepsilon_{yz}(u,v,X_r)-\bar{\phi}_y(u,X_t)\varepsilon_z(v,X_r)-\bar{\phi}_z(v,X_t)\varepsilon_y(u,X_r)\Big]^*\Big\}dW(u,v)K\left(\frac{X_r-X_t}{h}\right)K\left(\frac{X_s-X_t}{h}\right)\\
&+op(1)\\
\equiv\ & A_1+A_2+A_3+A_4+op(1).
\end{aligned}
$$

The result in Theorem 4.3.1 can be established given the following propositions.

Proposition C.2.1. *Under the conditions in Theorem 4.3.1,* $A_1 = op(1)$.

Proposition C.2.2. *Under the conditions in Theorem 4.3.1,* $A_2 = B + op(1)$, *where*

$$
\begin{aligned}
B &= h^{-k/2}\int\left\{\iint\left[1-|\phi_y(u,x)|^2-|\phi_z(v,x)|^2+|\phi_{yz}(u,v,x)|^2-2|\sigma(u,v)|^2\right]dW(u,v)\right\}a(x)dx\\
&\quad\times\int K^2(\tau)d\tau.
\end{aligned}
$$

Proposition C.2.3. *Under the conditions in Theorem 4.3.1,* $A_3 = op(1)$.

Proposition C.2.4. *Under the conditions in Theorem 4.3.1,* $A_4 = U + op(1)$ *and* $U/\sqrt{V} \xrightarrow{d} N(0,1)$, *where*

$$
\begin{aligned}
U &= \frac{2}{nh^{3k/2}}\sum_{1\le s<r\le n}\iiint\frac{a(x)}{f(x)}Re\Big\{\Big[(\varepsilon_{yz}(u,v,X_s)-\bar{\phi}_y(u,X_t)\varepsilon_z(v,X_s)-\bar{\phi}_z(v,X_t)\varepsilon_y(u,X_s)\Big]\\
&\quad\times\Big[\varepsilon_{yz}(u,v,X_r)-\bar{\phi}_y(u,X_t)\varepsilon_z(v,X_r)-\bar{\phi}_z(v,X_t)\varepsilon_y(u,X_r)\Big]^*\Big\}dW(u,v)K\left(\frac{X_s-x}{h}\right)K\left(\frac{X_r-x}{h}\right)dx,
\end{aligned}
$$

and

$$
V = 2\int\left[\iiiint\Lambda(u_1,u_2,v_1,v_2,x)dW(u_1,v_1)dW(u_2,v_2)\right]a^2(x)dx\int\left[\int K(\tau)K(\tau+\eta)d\tau\right]^2d\eta.
$$

Proposition C.2.5. *Under the conditions in Theorem 4.3.1,* $T_2 = op(1)$.

Proposition C.2.6. *Under the conditions in Theorem 4.3.1,* $T_3 = op(1)$.

□

Proof of Proposition C.1.1. Let

$$\eta_1(X_t) = \frac{a(X_t)}{f(X_t)^2} \iint \left|\varepsilon_{yz}(u,v,X_t) - \bar{\phi}_y(u,X_t)\varepsilon_z(v,X_t) - \bar{\phi}_z(v,X_t)\varepsilon_y(u,X_t)\right|^2 dW(u,v),$$

then we can write A_1 as

$$A_1 = \frac{1}{n^2 h^{3k/2}} K^2(0) \sum_{t=1}^{n} \eta_1(X_t).$$

Further we decompose A_1 as

$$\begin{aligned} A_1 &= \frac{1}{n^2 h^{3k/2}} K^2(0) \sum_{t=1}^{n} [\eta_1(X_t) - E\eta_1(X_t)] + \frac{1}{n h^{3k/2}} K^2(0)\, E\eta_1(X_t) \\ &\equiv A_{11} + A_{12}. \end{aligned}$$

Given Assumption 4.3.2 and Assumption 4.3.3, we have $A_{12} = O(n^{-1}h^{-3k/2}) = o(1)$. For A_{11}, it is easy to verify that $E(A_{11}) = 0$ and

$$\begin{aligned} var(A_{11}) &\leq K^4(0) n^{-4} h^{-3k} \sum_{t=1}^{n} Var[\eta_1(X_t)] + 2K^4(0) n^{-3} h^{-3k} \sum_{j=1}^{n-1} |Cov[\eta_1(X_1), \eta_1(X_{1+j})]| \\ &= O(n^{-3}h^{-3k}) + O(n^{-3}h^{-3k}) \sum_{j=1}^{n} j^2 \beta(j)^{\delta/(1+\delta)} \\ &= O(n^{-3}h^{-3k}), \end{aligned}$$

by the mixing condition. Thus we have $A_{11} = op(1)$ by Chebyshev's inequality. Combining $A_{12} = o(1)$ and $A_{11} = op(1)$, we complete the proof. □

Proof of Proposition C.1.2. Define

$$\begin{aligned} \eta_2(\xi_s, \xi_t) &= \iint \left|\varepsilon_{yz}(u,v,X_s) - \bar{\phi}_y(u,X_t)\varepsilon_z(v,X_s) - \bar{\phi}_z(v,X_t)\varepsilon_y(u,X_s)\right|^2 dW(u,v) \frac{a(X_t)}{f(X_t)^2} K^2\left(\frac{X_s - X_t}{h}\right) \\ &\quad + \iint \left|\varepsilon_{yz}(u,v,X_t) - \bar{\phi}_y(u,X_s)\varepsilon_z(v,X_t) - \bar{\phi}_z(v,X_s)\varepsilon_y(u,X_t)\right|^2 dW(u,v) \frac{a(X_t)}{f(X_t)^2} K^2\left(\frac{X_s - X_t}{h}\right), \end{aligned}$$

and $\eta_2(\xi_s) = \int \eta_2(\xi_s, \xi_t) dP(\xi_t)$. Then we can rewrite A_2 as

$$A_2 = \frac{1}{n^2 h^{3k/2}} \sum_{1 \leq s < t \leq n} \eta_2(\xi_s, \xi_t)$$

$$
\begin{aligned}
&= \frac{1}{n^2h^{3k/2}} \sum_{1\le s<t\le n} [\eta_2(\xi_s,\xi_t) - \eta_2(\xi_s) - \eta_2(\xi_t) + \eta_2] + \frac{n-1}{n^2h^{3k/2}} \sum_{t=1}^{n} [\eta_2(\xi_t) - \eta_2] + \frac{n-1}{2nh^{3k/2}}\eta_2 \\
&= A_{21} + A_{22} + A_{23}, \text{say.}
\end{aligned}
$$

By Lemma A(ii) of Hjellvik et al. (1998), we have

$$
\begin{aligned}
E|A_{21}|^2 &\le \frac{C}{n^2h^{3k}} \left[E\left|\eta_2(\xi_s,\xi_t) - \eta_2(\xi_s) - \eta_2(\xi_t) + \eta_2\right|^{2(1+\delta)}\right]^{1/(1+\delta)} \sum_{j=1}^{n} j^2\beta(j)^{\delta/(1+\delta)} \\
&= O(n^{-2}h^{-3k+k/(1+\delta)}).
\end{aligned}
$$

By Chebyshev's inequality, we have $A_{21} = op(1)$. Next, for A_{22}, we have

$$
\begin{aligned}
E|A_{22}|^2 &\le \frac{1}{n^2h^{3k}} \sum_{t=1}^{n} Var\left[\eta_2(\xi_t)\right] + \frac{2}{nh^{3k}} \sum_{j=1}^{n-1} \left|Cov\left[\eta_2(\xi_1), \eta_2(\xi_{j+1})\right]\right| \\
&\le O(n^{-1}h^{-k}) + \frac{2}{n^2h^{-k}} \sum_{j=1}^{n} j^2\beta(j)^{\delta/(1+\delta)} \\
&= O(n^{-1}h^{-k}),
\end{aligned}
$$

where we used the mixing condition in Assumption 4.3.1(a), and change of variable. Then by Chebyshev's inequality, we have $A_{22} = op(1)$. At last, for A_{23}, we have

$$
\begin{aligned}
A_{23} &= \frac{n-1}{nh^{3k/2}} \iiiint \frac{a(x_1)}{f(x_1)} K^2\left(\frac{x_2-x_1}{h}\right) E\Big\{\Big[\varepsilon_{yz}(u,v,x_2) - \bar{\phi}_y(u,x_1)\varepsilon_z(v,x_2) \\
&\quad -\bar{\phi}_z(v,x_1)\varepsilon_y(u,x_2)\Big]^2 \Big| x_2\Big\} f(x_2)dx_1dx_2dW(u,v) \\
&= \frac{n-1}{nh^{3k/2}} \iiiint \frac{a(x_1)}{f(x_1)} K^2\left(\frac{x_2-x_1}{h}\right)\Big(E\left[|\varepsilon_{yz}(u,v,x_2)|^2\Big|x_2\right] + E\left[|\bar{\phi}_y(u,x_1)|^2|\varepsilon_z(v,x_2)|^2\Big|x_2\right] \\
&\quad + E\left[|\bar{\phi}_z(v,x_1)|^2|\varepsilon_y(u,x_2)|^2|x_2\right] + 2E\left[Re\left(\bar{\phi}_y(u,x_1)\varepsilon_z(v,x_2)\bar{\phi}_z(v,x_1)^*\varepsilon_y(u,x_2)^*\right)|x_2\right] \\
&\quad -2E\left[Re\left(\varepsilon_{yz}(u,v,x_2)\bar{\phi}_y(u,x_1)^*\varepsilon_z(v,x_2)^*\right)|x_2\right] \\
&\quad -2E\left[Re\left(\varepsilon_{yz}(u,v,x_2)\bar{\phi}_z(v,x_1)^*\varepsilon_y(u,x_2)^*\right)|x_2\right]\Big) f(x_2)dx_1dx_2dW(u,v) \\
&= \frac{n-1}{nh^{3k/2}} \iiiint \frac{a(x_1)}{f(x_1)} K^2\left(\frac{x_2-x_1}{h}\right)\Big([1-|\phi_{yz}(u,v,x_2)|^2] + |\bar{\phi}_y(u,x_1)|^2(1-|\phi_z(v,x_2)|^2)
\end{aligned}
$$

$$+|\bar{\phi}_z(v,x_1)|^2(1-|\phi_y(u,x_2)|^2)+2Re[\bar{\phi}_y(u,x_1)\bar{\phi}_z(v,x_1)^*(1-\phi_z(v,x_2))(1-\phi_y(u,x_2))^*\Big]$$
$$-2Re\left[\left(1-\phi_{yz}(u,v,x_2)\right)\bar{\phi}_y(u,x_1)^*(1-\phi_z(v,x_2))^*\right]$$
$$-2Re\left[\left(1-\phi_{yz}(u,v,x_2)\right)\bar{\phi}_z(v,x_1)^*\left(1-\phi_y(u,x_2)\right)^*\right]\Big)\times f(x_2)dx_1dx_2dW(u,v)\,.$$

Recall the definition for $\bar{\phi}_{yz}(u,v,x)$, $\bar{\phi}_y(u,x)$, and $\bar{\phi}_z(v,x)$, it is easy to show that $\bar{\phi}_{yz}(u,v,x)-\phi_{yz}(u,v,x)=\frac{1}{2}h^2\nabla^2\phi_{yz}(u,v,x)C_K+o(h^2)$ uniformly in $(u,v)\in\mathbb{R}^{p+q}$ and in $x\in F_x$, where $\nabla^2\phi_{yz}(u,v,x)=\partial^2\phi_{yz}(u,v,x)/\partial^2x$ is the Laplacian of $\phi_{yz}(u,v,x)$. This result holds for $\bar{\phi}_y(u,x)$ and $\bar{\phi}_z(v,x)$. Then substituting $\bar{\phi}_{yz}(u,v,x)$, $\bar{\phi}_y(u,x)$ and $\bar{\phi}_z(v,x)$ by $\phi_{yz}(u,v,x)$, $\phi_y(u,x)$, and $\phi_z(v,x)$ only generates higher order terms and the result is asymptotically equivalent.

Also, let $\tau=\frac{x_2-x_1}{h}$, we have

$$\begin{aligned}
A_{23} &= \frac{n-1}{nh^{3k/2}}\iiiint\frac{a(x_1)}{f(x_1)}K^2(\tau)\Big([1-|\phi_{yz}(u,v,x_1+\tau h)|^2]+|\phi_y(u,x_1)|^2(1-|\phi_z(v,x_1+\tau h)|^2)\\
&\quad+2Re[\phi_y(u,x_1)\phi_z(v,\mathbf{1})^*(1-\phi_z(v,x_1+\tau h))(1-\phi_y(u,x_1+\tau h))^*\Big]\\
&\quad+|\phi_z(v,x_1)|^2(1-|\phi_y(u,x_1+\tau h)|^2)\\
&\quad-2Re\left[\left(1-\phi_{yz}(u,v,x_1+\tau h)\right)\phi_y(u,x_1)^*(1-\phi_z(v,x_1+\tau h))^*\right]\\
&\quad-2Re\left[\left(1-\phi_{yz}(u,v,x_1+\tau h)\right)\phi_z(v,x_1)^*\left(1-\phi_y(u,x_1+\tau h)\right)^*\right]\Big)\\
&\quad\times f(x_1+\tau h)dx_1d\tau dW(u,v)\,h^k+op(1)\\
&= \frac{n-1}{nh^{k/2}}\iiiint\frac{a(x_1)}{f(x_1)}K^2(\tau)\Big([1-|\phi_{yz}(u,v,x_1)|^2]+|\phi_y(u,x_1)|^2(1-|\phi_z(v,x_1)|^2)\\
&\quad+|\phi_z(v,x_1)|^2(1-|\phi_y(u,x_1)|^2)+2Re[\phi_y(u,x_1)\phi_z(v,x_1)^*(1-\phi_z(v,x_1))(1-\phi_y(u,x_1))^*\Big]\\
&\quad-2Re\left[\left(1-\phi_{yz}(u,v,x_1)\right)\phi_y(u,x_1)^*(1-\phi_z(v,x_1))^*\right]\\
&\quad-2Re\left[\left(1-\phi_{yz}(u,v,x_1)\right)\phi_z(v,x_1)^*\left(1-\phi_y(u,x_1)\right)^*\right]\Big)f(x_1)dx_1d\tau dW(u,v)+op(1)\\
&= \frac{n-1}{nh^{k/2}}\iiint a(x_1)(1-|\phi_{yz}(u,v,x_1)|^2-|\phi_y(u,x_1)|^2-|\phi_z(v,x_1)|^2-2|\phi_y(u,x_1)|^2|\phi_z(v,x_1)|^2\\
&\quad+4Re[\phi_{yz}(u,v,x_1)\phi_y(u,x_1)^*\phi_z(v,x_1)^*]dx_1dW(u,v)\int K^2(\tau)\,d\tau+op(1),
\end{aligned}$$

where for the second to last equality we use Assumption 4.3.1(c), Assumption

4.3.2, and second order Taylor expansion. Note that under the $\mathbb{H}_0$, we have

$$\sigma(u,v) = \sigma(u,v,x) = \phi_{yz}(u,v,x) - \phi_y(u,x)\phi_z(v,x),$$

then we can further simplify A_{23} as

$$\begin{aligned} A_{23} &= \frac{n-1}{nh^{k/2}} \iiint a(x_1)(1-|\phi_y(u,x_1)|^2 - |\phi_z(v,x_1)|^2 + |\phi_{yz}(u,v,x_1)|^2 - 2|\sigma(u,v)|^2)dx_1 dW(u,v) \\ &\quad \times \int K^2(\tau)\,d\tau + op(1) \\ &= B + op(1). \end{aligned}$$

and we finish the Proof of Proposition C.1.2. □

Proof of Proposition C.1.3. Define

$$\begin{aligned} \eta_3(\xi_s,\xi_t) &= \frac{a(X_t)}{f(X_t)^2}K(0)K\left(\frac{X_s-X_t}{h}\right)\iint Re\Big\{\Big[\varepsilon_{yz}(u,v,X_s) - \bar{\phi}_y(u,X_t)\varepsilon_z(v,X_s) \\ &\quad -\bar{\phi}_z(v,X_t)\varepsilon_y(u,X_s)\Big]\Big[\varepsilon_{yz}(u,v,X_t) - \bar{\phi}_y(u,X_t)\varepsilon_z(v,X_t) - \bar{\phi}_z(v,X_t)\varepsilon_y(u,X_t)\Big]^*\Big\}dW(u,v) \\ &\quad +\frac{a(X_s)}{f(X_s)^2}K(0)K\left(\frac{X_t-X_s}{h}\right)\iint Re\Big\{\Big[(\varepsilon_{yz}(u,v,X_t) - \bar{\phi}_y(u,X_s)\varepsilon_z(v,X_t) \\ &\quad -\bar{\phi}_z(v,X_s)\varepsilon_y(u,X_t)\Big]\Big[\varepsilon_{yz}(u,v,X_s) - \bar{\phi}_y(u,X_s)\varepsilon_z(v,X_s) - \bar{\phi}_z(v,X_s)\varepsilon_y(u,X_s))\Big]^*\Big\}dW(u,v), \end{aligned}$$

then we can rewrite A_3 as

$$A_3 = \frac{1}{n^2h^{3k/2}} \sum_{1<s\le t<n} \eta_3(\xi_s,\xi_t).$$

Also for any $\bar{\xi} \in \mathbb{R}^{p+q+k}$, we have $\int \eta_3(\xi_s,\bar{\xi})dP(\xi_s) = \int \eta_3(\bar{\xi},\xi_t)dP(\xi_t) = 0$. Thus, we can apply Lemma A(ii) of Hjellvik et al. (1998) and get

$$E|A_3|^2 \le \frac{C}{n^2h^{3k}}\left[E|\eta_3(\xi_s,\xi_t)|^{2(1+\delta)}\right]^{1/(1+\delta)} \sum_{j=1}^{n-1} j^2\beta(j)^{\delta/(1+\delta)} = O(n^{-2}h^{-3k+k/(1+\delta)}) = o(1).$$

Then the result in Proposition C.1.3 is established by Chebyshev's inequality. □

Proof of Proposition C.1.4. Define

$$\begin{aligned}
\eta_4(\xi_s,\xi_r,\xi_t) &= \frac{a(X_t)}{f(X_t)^2}K\left(\frac{X_s-X_t}{h}\right)K\left(\frac{X_r-X_t}{h}\right)\iint Re\Big\{\Big[\varepsilon_{yz}(u,v,X_s)-\bar{\phi}_y(u,X_t)\varepsilon_z(v,X_s)\\
&\quad -\bar{\phi}_z(v,X_t)\varepsilon_y(u,X_s)\Big]\Big[\varepsilon_{yz}(u,v,X_r)-\bar{\phi}_y(u,X_t)\varepsilon_z(v,X_r)-\bar{\phi}_z(v,X_t)\varepsilon_y(u,X_r)\Big]^*\Big\}dW(u,v)\\
&\quad +\frac{a(X_s)}{f(X_s)^2}K\left(\frac{X_t-X_s}{h}\right)K\left(\frac{X_r-X_s}{h}\right)\iint Re\Big\{\Big[\varepsilon_{yz}(u,v,X_t)-\bar{\phi}_y(u,X_s)\varepsilon_z(v,X_t)\\
&\quad -\bar{\phi}_z(v,X_s)\varepsilon_y(u,X_t)\Big]\Big[\varepsilon_{yz}(u,v,X_r)-\bar{\phi}_y(u,X_s)\varepsilon_z(v,X_r)-\bar{\phi}_z(v,X_s)\varepsilon_y(u,X_r)\Big]^*\Big\}dW(u,v)\\
&\quad +\frac{a(X_r)}{f(X_r)^2}K\left(\frac{X_s-X_r}{h}\right)K\left(\frac{X_t-X_r}{h}\right)\iint Re\Big\{\Big[\varepsilon_{yz}(u,v,X_s)-\bar{\phi}_y(u,X_r)\varepsilon_z(v,X_s)\\
&\quad -\bar{\phi}_z(v,X_r)\varepsilon_y(u,X_s)\Big]\Big[\varepsilon_{yz}(u,v,X_t)-\bar{\phi}_y(u,X_r)\varepsilon_z(v,X_t)-\bar{\phi}_z(v,X_r)\varepsilon_y(u,X_t)\Big]^*\Big\}dW(u,v),
\end{aligned}$$

and

$$\begin{aligned}
\eta_4(\xi_s,\xi_r) &= \int \eta_4(\xi_s,\xi_r,\xi)dP(\xi)\\
&= \iiint \frac{a(x)}{f(x)}K\left(\frac{X_s-x}{h}\right)K\left(\frac{X_r-x}{h}\right)Re\Big\{\Big[\varepsilon_{yz}(u,v,X_s)-\bar{\phi}_y(u,x)\varepsilon_z(v,X_s)\\
&\quad -\bar{\phi}_z(v,x)\varepsilon_y(u,X_s)\Big]\Big[\varepsilon_{yz}(u,v,X_r)-\bar{\phi}_y(u,x)\varepsilon_z(v,X_r)-\bar{\phi}_z(v,x)\varepsilon_y(u,X_r)\Big]^*\Big\}dW(u,v)dx.
\end{aligned}$$

Then, we decompose A_4 as

$$\begin{aligned}
A_4 &= \frac{1}{3n^2h^{3k/2}}\sum_{s\neq r\neq t}\left[\eta_4(\xi_s,\xi_r,\xi_t)-\eta_4(\xi_s,\xi_r)-\eta_4(\xi_s,\xi_t)-\eta_4(\xi_r,\xi_t)\right]\\
&\quad +\frac{1}{nh^{3k/2}}\sum_{s\neq r}\eta_4(\xi_s,\xi_r)\\
&= A_{41}+U+op(1),
\end{aligned}$$

where $op(1)$ is generated from replacing $\bar{\phi}$ by ϕ. Thus, we only need to show that $A_{41}=op(1)$. Note that

$$M=\int|\eta_4(\xi_s,\xi_r,\xi_t)|^2dP=O(h^{2k})$$

where P denotes any one of the following probability measures in the set

$$\{P(\xi_s,\xi_r,\xi_t),P(\xi_s)P(\xi_r,\xi_t),P(\xi_t)P(\xi_s,\xi_r),P(\xi_r)P(\xi_s,\xi_t),P(\xi_t)P(\xi_s)P(\xi_r)\}.$$

By Lemma A(i) of Hjellvik et al. (1998), we have

$$E|A_4|^2 \leq Cn^{-1}h^{-3k}M^{1/(1+\delta)}\sum_{j=1}^{\infty} j^2\beta(j)^{\delta/(1+\delta)} = O(n^{-1}h^{-3k+2k/(1+\delta)}).$$

Thus we have $A_4 = op(1)$.

Next, let's show $U/\sqrt{V} \xrightarrow{d} N(0,1)$, where $V = Var\,[U]$. Recall

$$\begin{aligned} U &= \frac{2}{nh^{3k/2}}\sum_{1\leq s<r\leq n}\iiint \frac{a(x)}{f(x)}K\left(\frac{X_s-x}{h}\right)K\left(\frac{X_r-x}{h}\right)Re\Big\{\Big[\varepsilon_{yz}(u,v,X_s)-\phi_y(u,x)\varepsilon_z(v,X_s) \\ &\quad -\bar{\phi}_z(v,x)\varepsilon_y(u,X_s)\Big]\Big[\varepsilon_{yz}(u,v,X_r)-\bar{\phi}_y(u,x)\varepsilon_z(v,X_r)-\bar{\phi}_z(v,x)\varepsilon_y(u,X_r)\Big]^*\Big\}dW(u,v)dx \\ &\equiv \frac{2}{nh^{3k/2}}\sum_{1\leq s<r\leq n} U(\xi_s,\xi_r), \end{aligned}$$

It is easy to verify that $E\,[U\,(\xi_s,\xi)] = E[U\,(\xi,\xi_r)] = 0$ for any $\xi \in \mathbb{R}^{p+q+k}$. Following Tenreiro's (1997) central limit theorem for degenerate U-statistics in a time series context, we can show $\sigma_n^{-1}\sum_{1\leq s<r\leq n} U(\xi_s,\xi_r) \xrightarrow{d} N(0,1)$ by showing the following conditions hold: For some constant $\delta_0 > 0, \gamma_0 < 1/2$ and $\gamma_1 > 0$, (i) $u_n(4+\delta_0) = O(n^{\gamma_0})$; (ii) $v_n(2) = o(1)$; (iii) $w_n(2+\delta_0/2) = o\left(n^{1/2}\right)$; and (iv) $z_n(2)n^{\gamma_1} = O(1)$, where $\sigma_n^2 = [Var\sum_{1\leq s<r\leq n} U(\xi_s,\xi_r)]$, where

$$\begin{aligned} u_n(p) &= \max\left\{\max_{1\leq i\leq n}\|U(\xi_i,\xi_1)\|_p, \left\|U(\xi_1,\bar{\xi}_1)\right\|_p\right\}, \\ v_n(p) &= \max\left\{\max_{1\leq i\leq n}\|G_{n1}(\xi_i,\xi_1)\|_p, \left\|G_{n1}(\xi_1,\bar{\xi}_1)\right\|_p\right\}, \\ w_n(p) &= \|G_{n1}(\xi_1,\xi_1)\|_p, \\ z_n(p) &= \max_{1\leq i\leq n}\max_{1\leq j\leq n}\left\{\left\|G_{nj}(\xi_i,\xi_1)\right\|_p, \left\|G_{nj}(\xi_1,\xi_i)\right\|_p, \left\|G_{nj}(\xi_1,\bar{\xi}_1)\right\|_p\right\} \end{aligned}$$

with $G_{ni}(\eta,\tau) = E\,[U(\xi_i,\eta)U(\xi_1,\tau)]$. Here we denote $\bar{\xi}$ as an independent copy of ξ, and denote $\|\cdot\|_p = \{E\,|\cdot|^p\}^{1/p}$ for $p \geq 1$.

To verify conditions (i) to (iv), we need first calculate the asymptotic variance of $U\,(\xi_s,\xi_r)$. However, the expression for it is very tedious which consists 36 terms and cannot be further simplified. For space, we write it in a compact form which is given as below.

$$
\begin{aligned}
var\left[U\left(\xi_s,\xi_r\right)\right] &= \iint \left|U\left(\xi_s,\xi_r\right)\right|^2 dP(\xi_s)P(\xi_r)\\
&= \iint \Bigg| \iiint \frac{a(x)}{f(x)} K\left(\frac{X_s-x}{h}\right) K\left(\frac{X_r-x}{h}\right) Re\Big\{\Big[\varepsilon_{yz}(u,v,X_s)-\phi_y(u,x)\varepsilon_z(v,X_s)\\
&\quad -\phi_z(v,x)\varepsilon_y(u,X_s)\Big]\Big[\varepsilon_{yz}(u,v,X_r)-\phi_y(u,x)\varepsilon_z(v,X_r)-\phi_z(v,x)\varepsilon_y(u,X_r)\Big]^*\Big\}\\
&\quad \times dW(u,v)dx\Bigg|^2 dP(\xi_s)P(\xi_r).
\end{aligned}
$$

Let $\tau=\frac{X_s-x}{h}$ and $\eta=\frac{X_s-X_r}{h}$, by change of variable

$$
\begin{aligned}
var[U\left(\xi_s,\xi_r\right)] &= \iint \Bigg| \iiint \frac{a(X_s-\tau h)}{f(X_s-\tau h)} K(\tau)K(\tau+\eta) Re\Big\{\Big[\varepsilon_{yz}(u,v,X_s)-\phi_y(u,X_s-\tau h)\varepsilon_z(v,X_s)\\
&\quad -\phi_z(v,X_s-\tau h)\varepsilon_y(u,X_s)\Big]\Big[\varepsilon_{yz}(u,v,X_r)-\phi_y(u,X_r-h\tau+\eta h)\varepsilon_z(v,X_r)\\
&\quad -\phi_z(v,X_r-h\tau+\eta h)\varepsilon_y(u,X_r)\Big]^*\Big\} dW(u,v)h^k d\tau\Bigg|^2 dP(\xi_s)P(\xi_r)\\
&= \iint \Bigg| \iiint \frac{a(X_s)}{f(X_s)} K(\tau)K(\tau+\eta) Re\Big\{\Big[\varepsilon_{yz}(u,v,X_s)-\phi_y(u,X_s)\varepsilon_z(v,X_s)\\
&\quad -\phi_z(v,X_s)\varepsilon_y(u,X_s)\Big]\Big[\varepsilon_{yz}(u,v,X_r)-\phi_y(u,X_r)\varepsilon_z(v,X_r)-\phi_z(v,X_r)\varepsilon_y(u,X_r)\Big]^*\Big\}\\
&\quad dW(u,v)h^k d\tau\Bigg|^2 dP(\xi_s)P(\xi_r)+op(1)\\
&= h^{2k}\iint \frac{a^2(X_s)}{f^2(X_s)}\Bigg| \iint Re\Big\{\Big[\varepsilon_{yz}(u,v,X_s)-\phi_y(u,X_s)\varepsilon_z(v,X_s)-\phi_z(v,X_s)\varepsilon_y(u,X_s)\Big]\\
&\quad \times\Big[\varepsilon_{yz}(u,v,X_r)-\phi_y(u,X_r)\varepsilon_z(v,X_r)-\phi_z(v,X_r)\varepsilon_y(u,X_r)\Big]^*\Big\} dW(u,v)\Bigg|^2 dP(\xi_s)P(\xi_r)\\
&\quad \times\left|\int K(\tau)K(\tau+\eta)d\tau\right|^2+op(1),
\end{aligned}
$$

where for the second to last inequality we use Assumption 4.3.1(c), Assumption 4.3.2 and the second order Taylor expansion. Further we have

$$
\begin{aligned}
var[U\left(\xi_s,\xi_r\right)] &= h^{2k}\iint \frac{a^2(X_s)}{f^2(X_s)}\Bigg|\Bigg\{ \iint Re\Big\{\Big[\varepsilon_{yz}(u_1,v_1,X_s)-\phi_y(u_1,X_s)\varepsilon_z(v_1,X_s)-\phi_z(v_1,X_s)\varepsilon_y(u_1,X_s)\Big]\\
&\quad \times\Big[\varepsilon_{yz}(u_1,v_1,X_r)-\phi_y(u_1,X_r)\varepsilon_z(v_1,X_r)-\phi_z(v_1,X_r)\varepsilon_y(u_1,X_r)\Big]^*\Big\} dW(u_1,v_1)\Bigg\}
\end{aligned}
$$

$$
\begin{aligned}
&\times\Bigg\{\iint Re\Big\{\Big[\varepsilon_{yz}(u_2,v_2,X_s)-\phi_y(u_2,X_s)\varepsilon_z(v_2,X_s)-\phi_z(v_2,X_s)\varepsilon_y(u_2,X_s)\Big]\\
&\times\Big[\varepsilon_{yz}(u_2,v_2,X_r)-\phi_y(u_2,X_r)\varepsilon_z(v_2,X_r)-\phi_z(v_2,X_r)\varepsilon_y(u_2,X_r)\Big]^*\Big\}dW(u_2,v_2)\Bigg\}\Bigg|^2\\
&\times dP(\xi_s)P(\xi_r)\left|\int K(\tau)K(\tau+\eta)d\tau\right|^2+op(1).
\end{aligned}
$$

Now, let's define the following, for $j\in\{1,2\}$:

$$
\begin{aligned}
\Phi_{yz}(u_1+u_2,v_1+v_2,x) &= \phi_{yz}(u_1+u_2,v_1+v_2,x)-\phi_{yz}(u_1,v_1,x)\phi_{yz}(u_1,v_1,x)\\
\Phi_{yz}(u_j,v_1+v_2,x) &= \phi_{yz}(u_j,v_1+v_2,x)-\phi_{yz}(u_j,v_1,x)\phi_z(v_2,x)\\
\Phi_{yz}(u_1+u_2,v_j,x) &= \phi_{yz}(u_1+u_2,v_j,x)-\phi_{yz}(u_1,v_j,x)\phi_y(u_2,x)\\
\Phi_y(u_1+u_2,x) &= \phi_y(u_1+u_2,x)-\phi_y(u_1,x)\phi_y(u_2,x)\\
\Phi_z(v_1+v_2,x) &= \phi_z(v_1+v_2,x)-\phi_z(v_1,x)\phi_z(v_2,x).
\end{aligned}
$$

Then by tedious calculation, we can show

$$
\begin{aligned}
var[U(\xi_s,\xi_r)] &= h^{3k}\int a^2(x)\iiiint \Lambda(u_1,u_2,v_1,v_2,x)dW(u_1,v_1)dW(u_2,v_2)dx\\
&\times\int\left|\int K(\tau)K(\tau+\eta)d\tau\right|^2 d\eta,
\end{aligned}
$$

where

$$
\begin{aligned}
\Lambda(u_1,u_2,v_1,v_2,x) &= |\Phi_{yz}(u_1+u_2,v_1+v_2,x)|^2+|\Phi_y(u_1+u_2,x)\phi_z(v_1,x)\phi_z(v_2,x)|^2\\
&+|\Phi_z(v_1+v_2,x)\phi_y(u_1,x)\phi_y(u_2,x)|^2+2|\sigma(u_2,v_1,x)\phi_y(u_1,x)\phi_z(v_2,x)|^2\\
&+2|\Phi_{yz}(u_1,v_1+v_2,x)\phi_y(u_2,x)|^2+2|\Phi_{yz}(u_1+u_2,v_1,x)\phi_z(v_2,x)|^2\\
&+4Re\left[\Phi_{yz}(u_1+u_2,v_1+v_2,x)\Phi_y(u_1+u_2,x)^*\phi_z(v_1,x)^*\phi_z(v_2,x)^*\right]\\
&+4Re\left[\Phi_{yz}(u_1+u_2,v_1+v_2,x)\Phi_z(v_1+v_2,x)^*\phi_y(u_1,x)^*\phi_y(u_2,x)^*\right]\\
&4Re\left[\Phi_y(u_1+u_2,x)\phi_z(v_1,x)\phi_z(v_2,x)\Phi_z(v_1+v_2,x)^*\phi_y(u_1,x)^*\phi_y(u_2,x)^*\right]\\
&4Re\left[\Phi_{yz}(u_1,v_1+v_2,x)\phi_y(u_2,x)\Phi_{yz}(u_1+u_2,v_1,x)^*\phi_z(v_2,x)^*\right]\\
&-4Re\left[\Phi_{yz}(u_1+u_2,v_1+v_2,x)\Phi_{yz}(u_1,v_1+v_2,x)^*\phi_y(u_2,x)^*\right]
\end{aligned}
$$

$$
\begin{aligned}
&-4Re\left[\Phi_{yz}(u_1+u_2,v_1+v_2,x)\Phi_{yz}(u_1+u_2,v_1,x)^*\phi_z(v_2,x)^*\right]\\
&-12Re\left[\Phi_{yz}(u_1,v_1+v_2,x)\phi_y(u_2,x)\sigma(u_2,v_1,x)^*\phi_y(u_1,x)^*\phi_z(v_2,x)^*\right]\\
&-4Re\left[\Phi_{yz}(u_1,v_1+v_2,x)\phi_y(u_2,x)\Phi_z(v_1+v_2,x)^*\phi_y(u_1,x)^*\phi_y(u_2,x)^*\right]\\
&-4Re\left[\Phi_{yz}(u_1+u_2,v_1,x)\phi_z(v_2,x)\sigma(u_2,v_1,x)^*\phi_y(u_1,x)^*\phi_z(v_2,x)^*\right]\\
&-4Re\left[\Phi_{yz}(u_1+u_2,v_1,x)\phi_z(v_2,x)\Phi_y(u_1+u_2,x)^*\phi_z(v_1,x)^*\phi_z(v_2,x)^*\right]\\
&+4Re\left[\sigma(u_2,v_1,x)\phi_y(u_1,x)\phi_z(v_2,x)\Phi_y(u_1+u_2,x)^*\phi_z(v_1,x)^*\phi_z(v_2,x)^*\right]\\
&+4Re\left[\sigma(u_2,v_1,x)\phi_y(u_1,x)\phi_z(v_2,x)\Phi_z(v_1+v_2,x)^*\phi_y(u_1,x)^*\phi_y(u_2,x)^*\right]\\
&+8Re\left[\Phi_{yz}(u_1+u_2,v_1+v_2,x)\sigma(u_2,v_1,x)^*\phi_y(u_1,x)^*\phi_z(v_2,x)^*\right]\\
&-8Re\left[\Phi_{yz}(u_1,v_1+v_2,x)\phi_y(u_2,x)\Phi_y(u_1+u_2,x)^*\phi_z(v_1,x)^*\phi_z(v_2,x)^*\right].
\end{aligned}
$$

Though the expression for the asymptotic variance looks messy, we only need its order of magnitude. For empirical application, we have provided a simple way of calculation and it is provided in Theorem 4.4.2. By Assumption 4.3.2 and Assumption 4.3.3, we know that $Var[U(\xi_s,\xi_r)] = O(h^{3k})$. Given $\sigma_n^2 = \sum_{1\le s<r\le n} Var[U(\xi_s,\xi_r)] = \frac{n^2}{2}Var[U(\xi_s,\xi_r)] + op(1)$ and $V = Var[U]$, we have

$$
\begin{aligned}
V &= \frac{4}{n^2h^{3k}}\frac{n^2}{2}Var[U(\xi_s,\xi_r)]\\
&= \frac{2}{h^{3k}}Var[U(\xi_s,\xi_r)]\\
&= 2\int a^2(x)\iiiint \Lambda(u_1,u_2,v_1,v_2,x)dW(u_1,v_1)dW(u_2,v_2)dx\int\left|\int K(\tau)K(\tau+\eta)d\tau\right|^2 d\eta.
\end{aligned}
$$

Next, let's verify condition (i) to (iv). Note that there will be 9 terms in $U(\xi_s,\xi_r)$ and each term has the same order of magnitude. To verify condition (i) to (iv) for $U(\xi_s,\xi_r)$, it is sufficient to verify for any one of the terms in $U(\xi_s,\xi_r)$. Define

$$
U_1(\xi_s,\xi_r) = \iiint \frac{a(x)}{f(x)}K\left(\frac{X_s-x}{h}\right)K\left(\frac{X_r-x}{h}\right)Re[\varepsilon_{yz}(u,v,X_s)\varepsilon_{yz}(u,v,X_r)^*]dW(u,v)dx,
$$

and it is easy to see that $U_1(\xi_s, \xi_r)$ is just the first term in $U(\xi_s, \xi_r)$. $U_1(\xi_s, \xi_r)$ is exactly the same term as in Wang and Hong (2012). Then the verification of condition (i) to (iv) is exactly the same as in Wang and Hong (2012). According to the Proof of Proposition 8 in Wang and Hong (2012), we have condition (i) to (iv) hold. Then we finish the proof of Proposition C.1.4. □

Proof of Proposition C.1.5.

$$T_2 = h^{k/2} \sum_{t=1}^{n} \iint |\hat{\sigma}(u, v) - \sigma(u, v)|^2 a(X_t) dW(u, v).$$

By definition of $\hat{\sigma}(u, v)$ and $\sigma(u, v)$, we know that they are not functions of X_t. Without of loss of generality, let $a(X_t) \equiv 1$. We have

$$T_2 = nh^{k/2} \iint |\hat{\sigma}(u, v) - \sigma(u, v)|^2 dW(u, v).$$

Also, since $\hat{\sigma}(u, v)$ is just sample average for $\sigma(u, v)$, we have $\hat{\sigma}(u, v) - \sigma(u, v) = Op(n^{-1/2})$. Then we have $T_2 = Op(h^{k/2}) = op(1)$ and we finish the Proof for Proposition C.1.5. □

Proof of Proposition C.1.6.

$$T_3 = h^{k/2} \sum_{t=1}^{n} \iint Re[(\hat{\sigma}(u, v, X_t) - \sigma(u, v, X_t))(\hat{\sigma}(u, v) - \sigma(u, v))^*] a(X_t) dW(u, v).$$

Since $a(X_t)$ doesn't change the order of T_3, without loss of generality we assume $a(X_t) = 1$. Also, under the condition of Theorem 1, the existence of integration over $(u, v) \in \mathbb{R}^{p+q}$ has no impact on order of T_3, we define T_3' as

$$T_3' = h^{k/2} \sum_{t=1}^{n} Re[(\hat{\sigma}(u, v, X_t) - \sigma(u, v, X_t))(\hat{\sigma}(u, v) - \sigma(u, v))^*],$$

and we know T_3 and T_3' have the same order of magnitude. For simplicity of notation, denote $\phi_{yz}(u, v, X_t) = \phi_{yz}(X_t)$, $\phi_y(u, X_t) = \phi_y(X_t)$, $\phi_z(v, X_t) = \phi_z(X_t)$ and

$\phi_{yz}(u,v) = \phi_{yz}$, $\phi_y(u) = \phi_y$, $\phi_z(v) = \phi_z$. Then we have

$$
\begin{aligned}
T_3' &= h^{k/2} \sum\nolimits_{t=1}^{n} Re\{[(\hat{\phi}_{yz}(X_t) - \phi_{yz}(X_t)) - \phi_y(X_t)(\hat{\phi}_z(X_t) - \phi_z(X_t)) \\
&\quad -\phi_z(X_t)(\hat{\phi}_y(X_t) - \phi_y(X_t)) - (\hat{\phi}_y(X_t) - \phi_y(X_t))(\hat{\phi}_z(X_t) - \phi_z(X_t))] \\
&\quad \times(\hat{\phi}_{yz} - \phi_{yz}) - \phi_y(\hat{\phi}_z - \phi_z) - \phi_z(\hat{\phi}_y - \phi_y) - (\hat{\phi}_y - \phi_y)(\hat{\phi}_z - \phi_z)]^*\}dW(u,v) \\
&= h^{k/2} \sum\nolimits_{t=1}^{n} Re\{[\hat{\phi}_{yz}(X_t) - \phi_{yz}(X_t)](\hat{\phi}_{yz} - \phi_{yz})^* - \phi_y(X_t)[\hat{\phi}_z(X_t) - \phi_z(X_t)](\hat{\phi}_{yz} - \phi_{yz})^* \\
&\quad -\phi_z(X_t)[\hat{\phi}_y(X_t) - \phi_y(X_t)](\hat{\phi}_{yz} - \phi_{yz})^* - [\hat{\phi}_{yz}(X_t) - \phi_{yz}(X_t)]\phi_y^*(\hat{\phi}_z - \phi_z)^* \\
&\quad +\phi_y(X_t)[\hat{\phi}_z(X_t) - \phi_z(X_t)]\phi_y^*(\hat{\phi}_z - \phi_z)^* + \phi_z(X_t)[\hat{\phi}_y(X_t) - \phi_y(X_t)]\phi_y^*(\hat{\phi}_z - \phi_z)^* \\
&\quad -[\hat{\phi}_{yz}(X_t) - \phi_{yz}(X_t)]\phi_z^*(\hat{\phi}_y - \phi_y)^* + \phi_y(X_t)[\hat{\phi}_z(X_t) - \phi_z(X_t)]\phi_z^*(\hat{\phi}_y - \phi_y)^* \\
&\quad +\phi_z(X_t)[\hat{\phi}_y(X_t) - \phi_y(X_t)]\phi_z^*(\hat{\phi}_y - \phi_y)^* \\
&\quad -[\hat{\phi}_y(X_t) - \phi_y(X_t))(\hat{\phi}_z(X_t) - \phi_z(X_t)](\hat{\phi}_{yz} - \phi_{yz})^* \\
&\quad +[\hat{\phi}_y(X_t) - \phi_y(X_t))(\hat{\phi}_z(X_t) - \phi_z(X_t)]\phi_y^*(\hat{\phi}_z - \phi_z)^* \\
&\quad +[\hat{\phi}_y(X_t) - \phi_y(X_t))(\hat{\phi}_z(X_t) - \phi_z(X_t)]\phi_z^*(\hat{\phi}_y - \phi_y)^* \\
&\quad +[\hat{\phi}_y(X_t) - \phi_y(X_t))(\hat{\phi}_z(X_t) - \phi_z(X_t)](\hat{\phi}_y - \phi_y)^*(\hat{\phi}_z - \phi_z)^* \\
&\quad -[\hat{\phi}_{yz}(X_t) - \phi_{yz}(X_t)](\hat{\phi}_y - \phi_y)^*(\hat{\phi}_z - \phi_z)^* \\
&\quad +\phi_y(X_t)[\hat{\phi}_z(X_t) - \phi_z(X_t)](\hat{\phi}_y - \phi_y)^*(\hat{\phi}_z - \phi_z)^* \\
&\quad +\phi_z(X_t)[\hat{\phi}_y(X_t) - \phi_y(X_t)](\hat{\phi}_y - \phi_y)^*(\hat{\phi}_z - \phi_z)^*\} \\
&\equiv \sum_{i=1}^{16} S_i.
\end{aligned}
$$

We can show that S_1 to S_9 are of the same order and S_{10} to S_{16} are higher order terms. To show $T_3' = op(1)$, it is sufficient to show $S_1 = op(1)$.

$$
\begin{aligned}
S_1 &= h^{k/2} \sum\nolimits_{t=1}^{n} Re\{[\hat{\phi}_{yz}(X_t) - \phi_{yz}(X_t)](\hat{\phi}_{yz} - \phi_{yz})^*\} \\
&= h^{k/2} \sum\nolimits_{t=1}^{n} Re\{[\hat{\phi}_{yz}(X_t) - \bar{\phi}_{yz}(X_t) + \bar{\phi}_{yz}(X_t) - \phi_{yz}(X_t)](\hat{\phi}_{yz} - \phi_{yz})^*\} \\
&= h^{k/2} \sum\nolimits_{t=1}^{n} Re\{[\hat{\phi}_{yz}(X_t) - \bar{\phi}_{yz}(X_t)](\hat{\phi}_{yz} - \phi_{yz})^* \\
&\quad +h^{k/2} \sum\nolimits_{t=1}^{n} Re\{[\bar{\phi}_{yz}(X_t) - \phi_{yz}(X_t)](\hat{\phi}_{yz} - \phi_{yz})^*
\end{aligned}
$$

$$= \quad S_{11} + S_{12}.$$

Let's first look at S_{11}.

$$\begin{aligned}
S_{11} &= h^{k/2} \sum_{t=1}^{n} Re\{[\hat{\phi}_{yz}(X_t) - \bar{\phi}_{yz}(X_t)](\hat{\phi}_{yz} - \phi_{yz})^*\} \\
&= h^{k/2} \sum_{t=1}^{n} Re\left[\sum_{s=1}^{n} \frac{K\left(\frac{X_s - X_t}{h}\right)}{nh^k f(X_t)} \varepsilon_{yz}(u, v, X_s) \times \frac{1}{n} \sum_{s=1}^{n} \varepsilon_{yz}^{(s)}(u, v)^*\right] + op(1) \\
&= n^{-2} h^{-k/2} \sum_{t=1}^{n} \frac{1}{f(X_t)} \sum_{s=1}^{n} Re\left[K\left(\frac{X_s - X_t}{h}\right) \varepsilon_{yz}(u, v, X_s) \varepsilon_{yz}^{(s)}(u, v)^*\right] + op(1) \\
&= n^{-2} h^{-k/2} \sum_{t=1}^{n} \frac{K(0)}{f(X_t)} Re\left[\varepsilon_{yz}(u, v, X_t) \varepsilon_{yz}^{(t)}(u, v)^*\right] \\
&\quad + n^{-2} h^{-k/2} \sum_{s \neq t}^{n} \frac{1}{f(X_t)} K\left(\frac{X_s - X_t}{h}\right) Re\left[\varepsilon_{yz}(u, v, X_s) \varepsilon_{yz}^{(s)}(u, v)^*\right] + op(1) \\
&= S_{111} + S_{112} + op(1),
\end{aligned}$$

where $\varepsilon_{yz}^{(s)}(u, v) = e^{iuY_s + ivZ_s} - \phi_{yz}(u, v)$. Define $\varphi_1(X_t) = \frac{1}{f(X_t)} Re\left[\varepsilon_{yz}(u, v, X_t)\varepsilon_{yz}^{(t)}(u, v)^*\right]$, then we can write S_{111} as

$$\begin{aligned}
S_{111} &= n^{-2} h^{-k/2} K(0) \sum_{t=1}^{n} \varphi_1(X_t) \\
&= n^{-2} h^{-k/2} K(0) \sum_{t=1}^{n} [\varphi_1(X_t) - E\varphi_1(X_t)] + n^{-2} h^{-k/2} K(0) \sum_{t=1}^{n} E\varphi_1(X_t) \\
&= S_{1111} + S_{1112}.
\end{aligned}$$

Given $E(S_{1111}) = 0$ and

$$\begin{aligned}
var(S_{1111}) &\leq K^2(0) n^{-4} h^{-k} \sum_{t=1}^{n} Var[\varphi(X_t)] + 2K^2(0) n^{-3} h^{-k} \sum_{j=1}^{n-1} Cov\left|\left[\varphi(X_1), \varphi(X_{j+1})\right]\right| \\
&\leq O\left(n^{-3} h^{-k}\right) + O\left(n^{-3} h^{-k}\right) \sum_{j=1}^{\infty} j^2 \beta(j)^{\delta/(1+\delta)} \\
&= O\left(n^{-3} h^{-k}\right),
\end{aligned}$$

we can show $S_{1111} = op(1)$ by Chebyshev's inequality. And the constant $S_{1112} = O(n^{-2} h^{-k/2}) = o(1)$. Then we know that $S_{111} = op(1)$.

For S_{112}, define

$$\begin{aligned}\eta_5(\xi_s,\xi_t) &= \frac{1}{f(X_t)}K\left(\frac{X_s-X_t}{h}\right)Re\left[\varepsilon_{yz}(u,v,X_s)\varepsilon_{yz}^{(s)}(u,v)^*\right]\\ &\quad+\frac{1}{f(X_s)}K\left(\frac{X_t-X_s}{h}\right)Re\left[\varepsilon_{yz}(u,v,X_t)\varepsilon_{yz}^{(t)}(u,v)^*\right],\end{aligned}$$

$\eta_5(\xi_s)=\int\eta_5(\xi_s,\xi_t)dP(\xi_t)$, and $\eta_5=\int\eta_5(\xi_s)dP(\xi_s)$. Then we can rewrite S_{112} as

$$\begin{aligned}S_{112} &= n^{-2}h^{-k/2}\sum_{1\le s<t\le n}\eta_5(\xi_s,\xi_t)\\ &= n^{-2}h^{-k/2}\sum_{1\le s<t\le n}[\eta_5(\xi_s,\xi_t)-\eta_5(\xi_s)-\eta_5(\xi_t)+\eta_5]+\frac{n-1}{n^2h^{k/2}}\sum_{t=1}^{n}[\eta_5(\xi_t)-\eta_5]+\frac{n-1}{nh^{k/2}}\eta_5\\ &= S_{1121}+S_{1122}+S_{1123},\ \text{say.}\end{aligned}$$

By Lemma A(ii) of Hjellvik et al. (1998), we have

$$\begin{aligned}E\left|S_{1121}\right|^2 &\le \frac{C}{n^2h^k}\left[E\left|\eta_5(\xi_s,\xi_t)-\eta_5(\xi_s)-\eta_5(\xi_t)+\eta_5\right|^{2(1+\delta)}\right]^{1/(1+\delta)}\sum_{j=1}^{\infty}j^2\beta(j)^{\delta/(1+\delta)}\\ &= O(n^{-2}h^{-k+k/(1+\delta)}).\end{aligned}$$

Then we have $S_{1121}=op(1)$ by Chebyshev's inequality. Next, for S_{1122},

$$\begin{aligned}E\left|S_{1122}\right|^2 &\le \frac{1}{n^2h^k}\sum_{t=1}^{n}Var\left[\eta_5(\xi_t)\right]+\frac{2}{nh^k}\sum_{j=1}^{n-1}Cov\left|\left[\eta_5(\xi_1),\eta_5(\xi_{1+j})\right]\right|\\ &\le Op(n^{-1}h^k)+\frac{2h^k}{n}\sum_{j=1}^{\infty}j^2\beta(j)^{\delta/(1+\delta)}\\ &= Op(n^{-1}h^k),\end{aligned}$$

where we used the mixing condition, change of variable and Assumption 4.3.1. Thus $S_{1122}=op(1)$ by Chebyshev's inequality. Finally, by change of variable it is easy to show that

$$S_{1123}=Op(h^{k/2})=op(1).$$

Combining the above results, we show that $S_{112}=op(1)$ and thus $S_{11}=op(1)$.

At last, we decompose S_{12}.

$$S_{12} = h^{k/2}\sum_{t=1}^{n}Re\{[\bar{\phi}_{yz}(X_t)-\phi_{yz}(X_t)](\hat{\phi}_{yz}-\phi_{yz})^*\}$$

$$
\begin{aligned}
&= h^{k/2}\sum_{t=1}^{n} Re\{[\bar{\phi}_{yz}(X_t)-\phi_{yz}(X_t)]\frac{1}{n}\sum_{s=1}^{n}\varepsilon_{yz}^{(s)}(u,v)^*\}.\\
&= n^{-1}h^{k/2}\sum_{t=1}^{n} Re\left\{[\bar{\phi}_{yz}(X_t)-\phi_{yz}(X_t)]\varepsilon_{yz}^{(t)}(u,v)^*\right\}\\
&\quad +n^{-1}h^{k/2}\sum_{s\neq t}^{n} Re\left\{[\bar{\phi}_{yz}(X_t)-\phi_{yz}(X_t)]\varepsilon_{yz}^{(s)}(u,v)^*\right\}\\
&= S_{121}+S_{122.}
\end{aligned}
$$

By Assumption 4.3.3 and Assumption 4.3.4 we have $E(S_{121}) = 0$, and $var(S_{121}) = O(n^{-1}h^{k+4}) = o(1)$, where we use the result that $\bar{\phi}_{yz}(u,v,x) - \phi_{yz}(u,v,x) = \frac{1}{2}h^2\nabla^2\phi_{yz}(u,v,x)C_K + o(h^2)$ uniformly in $(u,v) \in \mathbb{R}^{p+q}$ and in $x \in F_x$, and $\nabla^2\phi_{yz}(u,v,x) = \partial^2\phi_{yz}(u,v,x)/\partial^2 x$ is the Laplacian of $\phi_{yz}(u,v,x)$. By Chebyshev's inequality we have $S_{121} = op(1)$.

By the fact that $E\left[\varepsilon_{yz}^{(t)}(u,v)\right] = 0$, we define

$$
\eta_6(\xi_s,\xi_t) = Re\left\{[\bar{\phi}_{yz}(X_t)-\phi_{yz}(X_t)]\varepsilon_{yz}^{(s)}(u,v)^*\right\} + Re\left\{[\bar{\phi}_{yz}(X_s)-\phi_{yz}(X_t)]\varepsilon_{yz}^{(t)}(u,v)^*\right\},
$$

and

$$
\eta_6(\xi_s) = \int \eta_5(\xi_s,\xi_t)dP(\xi_t) = \int Re\left\{[\bar{\phi}_{yz}(x)-\phi_{yz}(x)]\varepsilon_{yz}^{(s)}(u,v)^*\right\}dx.
$$

Then we can rewrite S_{122} as

$$
\begin{aligned}
S_{122} &= n^{-1}h^{k/2}\sum_{1\le s<t\le n}\eta_6(\xi_s,\xi_t)\\
&= n^{-1}h^{k/2}\sum_{1\le s<t\le n}[\eta_6(\xi_s,\xi_t)-\eta_6(\xi_s)-\eta_6(\xi_t)] + \frac{2(n-1)h^{k/2}}{n}\sum_{t=1}^{n}\eta_6(\xi_t)\\
&= S_{1221}+S_{1222}.
\end{aligned}
$$

By Lemma A(ii) of Hjellvik et al. (1998)

$$
\begin{aligned}
var(S_{1221}) &\le \frac{Ch^k}{n^2}n^2E\left[|\eta_6(\xi_s,\xi_t)-\eta_6(\xi_s)-\eta_6(\xi_t)|^{2(1+\delta)}\right]^{1/(1+\delta)}\sum_{j=1}^{\infty}j^2\beta(j)^{\delta/(1+\delta)}\\
&= O(h^{k+4/(1+\delta)}),
\end{aligned}
$$

and get $S_{1221} = op(1)$ by Chebyshev's inequality. Furthermore we show

$$\begin{aligned} Var(S_{1222}) &\leq \frac{4(n-1)^2h^k}{n^2}\sum_{t=1}^{n} Var\left[\eta_6(\xi_t)\right] + \frac{4(n-1)^2h^k}{n}\sum_{j=1}^{n-1} Cov\left|\left[\eta_6(\xi_1), \eta_6(\xi_{1+j})\right]\right| \\ &\leq O(nh^{k+4}) + O(nh^{k+4})\sum_{j=1}^{\infty} j^2\beta(j)^{\delta/(1+\delta)} \\ &= O(nh^{k+4}), \end{aligned}$$

by $\bar{\phi}_{yz}(u, v, x) - \phi_{yz}(u, v, x) = Op(h^2)$ and the mixing inequality. Thus we have $S_{1222} = op(1)$ by Chebyshev's inequality. Combining the above results, we finish the Proof of Proposition C.1.6. □

Proof of Theorem 4.3.2. According to the Proof of Theorem 1, the asymptotic distribution of our test statistic is only determined by the nonparametric estimator for condition generalized covariance function which is T_1 in Eq.(). This leading term coincides with the test statistic proposed in Wang and Hong (2012) and our test statistic and their test statistic both following normal distribution with different asymptotic mean and asymptotic variance. In spite of the difference in distribution parameters, the rate of local alternative that can be detected by our test statistic will be the same as in theirs. Thus, the Proof of Theorem 2 will be quite similar to the Proof of Theorem 2 in Wang and Hong (2012). For space, we neglect it. □

Proof of Theorem 4.4.1. First, let's look at the population version of the test statistic $T = \iiint \frac{1}{\sigma}(u, v, x) - \sigma(u, v)|^2 a(x) f(x) dW(u, v)dx$. We need to show

$$I_1 \equiv I_1(x) = \iint |\sigma(u, v, x) - \sigma(u, v)|^2 dW(u, v) < \infty.$$

By triangle inequality,

$$I_1 \leq 2\iint |\sigma(u, v, x)|^2 dW(u, v) + 2\iint |\sigma(u, v)^2 dW(u, v)$$

$$= \quad 2I_{11} + 2I_{12}.$$

By the definition of $\sigma(u, v, x)$,

$$\begin{aligned}
I_{11} &= \iint |E\{[e^{iuY} - \phi_y(u, x)]\left[e^{ivZ} - \phi_z(v, x)\right]|X = x\}|^2 dW(u, v) \\
&\leq \iint E\left\{[e^{iuY} - \phi_y(u, x)]^2|X = x\right\} E\left\{[e^{ivZ} - \phi_z(v, x)]^2|X = x\right\} dW(u, v) \\
&= \iint [1 - |\phi_y(u, x)|^2][1 - |\phi_z(v, x)|^2]dW(u, v),
\end{aligned}$$

by CauchySchwarz inequality. Let $\bar{Y}$ and $\bar{Z}$ be independent copy of Y and Z respectively, we have

$$\begin{aligned}
I_{11} &\leq \iint [1 - |\phi_y(u, x)|^2][1 - |\phi_z(v, x)|^2]dW(u, v) \\
&= \iint [1 - E(e^{iuY}|X = x)E(e^{-iu\bar{Y}}|X = x)] \\
&= \times[1 - E(e^{ivZ}|X = x)E(e^{-iv\bar{Z}}|X = x)]dW(u, v) \\
&= \iint [1 - E(\cos uY + i \sin uY|X = x)E(\cos u\bar{Y} - i \sin u\bar{Y}|X = x)] \\
&\quad \times[1 - E(\cos vZ + i \cos vZ|X = x)E(\cos v\bar{Z} + i \sin v\bar{Z}|X = x)]dW(u, v) \\
&= \iint [1 - E(\cos uY \cos u\bar{Y} + \sin uY \sin u\bar{Y}|X = x)] \\
&\quad \times[1 - E(\cos vZ \cos v\bar{Z} + \sin vZ \sin v\bar{Z}|X = x)]dW(u, v) \\
&= \iint [1 - E(\cos u(Y - \bar{Y})|X = x)][1 - E(\cos v(Z - \bar{Z})|X = x)]dW(u, v).
\end{aligned}$$

Given $dW(u, v) = \frac{1}{c_p|u|_p^{p+1}} \frac{1}{c_q|v|_q^{q+1}} dudv$, where $c_p = \frac{\pi^{(1+p)/2}}{\Gamma((1+p)/2)}$ and $c_q = \frac{\pi^{(1+q)/2}}{\Gamma((1+q)/2)}$ and apply the Lemma (Székely et al. 2007), we have

$$\begin{aligned}
I_{11} &\leq \iint [1 - E(\cos u(Y - \bar{Y})|X = x)][1 - E(\cos v(Z - \bar{Z})|X = x)]dW(u, v) \\
&= \iint E[(1 - \cos u(Y - \bar{Y}))|X = x]E[(1 - (\cos v(Z - \bar{Z}))|X = x]dW(u, v) \\
&= E\left(\left[\int \frac{1 - \cos u(Y - \bar{Y})}{c_p|u|_p^{p+1}} du\right]\middle| X = x\right) \cdot E\left(\left[\int \frac{1 - \cos v(Z - \bar{Z})}{c_q|v|_q^{q+1}} dv\right]\middle| X = x\right) \\
&= E\left(\left|Y - \bar{Y}\right|_p |X = x\right) \cdot E\left(\left|Z - \bar{Z}\right|_q |X = x\right) \\
&\leq \infty,
\end{aligned}$$

where the last inequality follows from Assumption 4.3.1(b) that $E\left(|Y|_p\,|X=x\right) < \infty$ and $E\left(|Z|_q\,|X=x\right) < \infty$. Now we have shown that $I_{11} < \infty$ under the conditions Theorem 4.4.1.

For I_{12}, by the definition of $\sigma(u,v)$,

$$\begin{aligned} I_{12} &= \iint |E\{[e^{iuY} - \phi_y(u)]\left[e^{ivZ} - \phi_z(v)\right]\}|^2 dW(u,v) \\ &\leq \iint E|e^{iuY} - \phi_y(u)|^2 E|e^{ivZ} - \phi_z(v,x)|^2 dW(u,v) \\ &= \iint [1 - |\phi_y(u)|^2][1 - |\phi_z(v)|^2] dW(u,v), \end{aligned}$$

where we use CauchySchwarz inequality. Let $\bar{Y}$ and $\bar{Z}$ be independent copy of Y and Z respectively, we have

$$\begin{aligned} I_{12} &\leq \iint [1 - |\phi_y(u)|^2][1 - |\phi_z(v)|^2] dW(u,v) \\ &= \iint [1 - E(e^{iuY})E(e^{-iu\bar{Y}})][1 - E(e^{ivZ})E(e^{-iv\bar{Z}})] dW(u,v) \\ &= \iint [1 - E(\cos uY + i \sin uY)E(\cos u\bar{Y} - i \sin u\bar{Y})] \\ &\quad \times [1 - E(\cos vZ + i \cos vZ)E(\cos v\bar{Z} + i \sin v\bar{Z})] dW(u,v) \\ &= \iint [1 - E(\cos uY \cos u\bar{Y} + \sin uY \sin u\bar{Y})] \\ &\quad \times [1 - E(\cos vZ \cos v\bar{Z} + \sin vZ \sin v\bar{Z})] dW(u,v) \\ &= \iint [1 - E(\cos u(Y - \bar{Y}))][1 - E(\cos v(Z - \bar{Z}))] dW(u,v) \end{aligned}$$

Applying the special weighting function that $dW(u,v) = \frac{1}{c_p|u|_p^{p+1}} \frac{1}{c_q|v|_q^{q+1}} dudv$, where $c_p = \frac{\pi^{(1+p)/2}}{\Gamma((1+p)/2)}$ and $c_q = \frac{\pi^{(1+q)/2}}{\Gamma((1+q)/2)}$ and by Lemma (Székely et al. 2007),

$$\begin{aligned} I_{12} &\leq \iint [1 - E(\cos u(Y - \bar{Y}))][1 - E(\cos v(Z - \bar{Z}))] dW(u,v) \\ &= \iint E[(1 - \cos u(Y - \bar{Y}))] E[(1 - (\cos v(Z - \bar{Z}))] dW(u,v) \\ &= E\left(\left[\int \frac{1 - \cos u(Y - \bar{Y})}{c_p|u|_p^{p+1}} du\right]\right) \cdot E\left(\left[\int \frac{1 - \cos v(Z - \bar{Z})}{c_q|v|_q^{q+1}} dv\right]\right) \\ &= E\left(\left|Y - \bar{Y}\right|_p\right) \cdot E\left(\left|Z - \bar{Z}\right|_q\right) \end{aligned}$$

$$< \quad \infty$$

where the last inequality follows from Assumption 4.3.1(b) that $E\left(|Y|_p\right) < \infty$ and $E\left(|Z|_q\right) < \infty$. Combining the above results, we have shown $I_1 < \infty$.

Next, let's deal with asymptotic mean. Given

$$\begin{aligned} B &= h^{-k/2} \int \left[\iint [1 - |\phi_y(u,x)|^2 - |\phi_z(v,x)|^2 + |\phi_{yz}(u,v,x)|^2 - 2|\sigma(u,v)|^2] dW(u,v) \right] a(x)dx \\ &\quad \times \int K^2(\tau)d\tau, \end{aligned}$$

we need to show that

$$I_2 \equiv I_2(x) = \iint [1 - |\phi_y(u,x)|^2 - |\phi_z(v,x)|^2 + |\phi_{yz}(u,v,x)|^2 - 2|\sigma(u,v)|^2] dW(u,v) < \infty.$$

Under $\mathbb{H}_0$, We decompose

$$\begin{aligned} I_2 &= \iint [1 - |\phi_y(u,x)|^2 - |\phi_z(v,x)|^2 + 1 - 1 + |\phi_{yz}(u,v,x)|^2 - 2|\sigma(u,v)|^2] dW(u,v) \\ &= \iint (1 - |\phi_y(u,x)|^2) dW(u,v) - \iint (1 - |\phi_z(u,x)|^2) dW(u,v) \\ &\quad - \iint (1 - |\phi_{yz}(u,v,x)|^2) dW(u,v) - 2 \iint |\sigma(u,v)|^2 dW(u,v) \\ &= I_{21} - I_{22} - I_{23} - 2I_{24}. \end{aligned}$$

We immediately have $I_{21} < \infty$, $I_{22} < \infty$ following from the result that $I_{11} < \infty$ and $I_{24} < \infty$ since $I_{24} = I_{12}$. We only need to deal with I_{23}. Let $(\bar{Y}', \bar{Z}')'$ be an independent copy of $(Y', Z')'$, then we have

$$\begin{aligned} I_{23} &= \iint (1 - |\phi_{yz}(u,v,x)|^2) dW(u,v) \\ &= \iint [1 - E\left(e^{iuY+ivZ}|X=x\right) E\left(e^{iu\bar{Y}+iv\bar{Z}}|X=x\right)^*] dW(u,v) \\ &= \iint \{1 - E[\cos(uY+vZ) + i\sin(uY+vZ)|X=x] \\ &= \times E[\cos\left(u\bar{Y}+v\bar{Z}\right) - i\sin(u\bar{Y}+v\bar{Z})|X=x]\} dW(u,v) \\ &= \iint \{1 - E\left[\cos(uY+vZ)\cos\left(u\bar{Y}+v\bar{Z}\right) + \sin(uY+vZ)\sin(u\bar{Y}+v\bar{Z})|X=x\right]\} dW(u,v) \end{aligned}$$

$$
\begin{aligned}
&= \iint \{1 - E(\cos[u(Y-\bar{Y}) + v(Z-\bar{Z})]|X=x)\} dW(u,v) \\
&= \iint \{1 - E(\cos\left[u(Y-\bar{Y})\right]\cos[v(Z-\bar{Z})]|X=x)\} dW(u,v) \\
&\quad + \iint E(\sin\left[u(Y-\bar{Y})\right]\sin[v(Z-\bar{Z})]|X=x\right]) dW(u,v) \\
&= \iint \{1 - E(\cos\left[u(Y-\bar{Y})\right]\cos[v(Z-\bar{Z})]|X=x\right])\} dW(u,v),
\end{aligned}
$$

where we claim that $\iint E(\sin\left[u(Y-\bar{Y})\right]\sin[v(Z-\bar{Z})]|X=x\right])dW(u,v) = 0$ since sin function is an odd function and the weighting function $dW(u,v)$ is symmetric about the origin. By the following equality

$$
\cos\alpha\cos\beta = 1 - (1-\cos\alpha) - (1-\cos\beta) + (1-\cos\alpha)(1-\cos\beta)
$$

we have

$$
\begin{aligned}
I_{23} &= \iint E\Big\{\left(1-\cos\left[u(Y-\bar{Y})\right]\right) + \left(1-\cos\left[v(Z-\bar{Z})\right]\right) \\
&\quad - \left(1-\cos\left[u(Y-\bar{Y})\right]\right)\left(1-\cos\left[v(Z-\bar{Z})\right]\right)|X=x\Big\} dW(u,v).
\end{aligned}
$$

Applying the special weighting function and we get

$$
I_{23} = E\left(\left|Y-\bar{Y}\right|_p |X=x\right) + E\left(\left|Z-\bar{Z}\right|_q |X=x\right) - E\left(\left|Y-\bar{Y}\right|_p \left|Z-\bar{Z}\right|_q |X=x\right).
$$

Then by Assumption 4.3.1(b), we have $I_{23} < \infty$. Combining the results we have proved that $I_2 < \infty$.

At last, let's show $V < \infty$. Recall

$$
V = 2\int a^2(x) \iiiint \Lambda(u_1,u_2,v_1,v_2,x)dW(u_1,v_1)dW(u_2,v_2)dx \int \left|\int K(\tau)K(\tau+\eta)d\tau\right|^2 d\eta.
$$

Define $I_3 \equiv I_3(x) = \iiiint \Lambda(u_1,u_2,v_1,v_2,x)dW(u_1,v_1)dW(u_2,v_2)$ and we need to show $I_3 < \infty$. Since the expression for $\Lambda(u_1,u_2,v_1,v_2,x)$ contains many terms and we need to show each of them are bounded under the integration using the special weighting function. Intuitively, this must hold because we can regard the asymptotic variance as the square of remaining term after subtracting asymptotic mean from the test statistic which are both bounded. If V is not bounded, then at least one of the test statistic and the asymptotic has to be unbounded and this contradicts the results we obtained above. It can be shown, though under tedious derivation, that $I_3 < \infty$. For space, we neglect it. □

Proof of Theorem 4.4.2. We first show $h^{k/2}\sum_{t=1}^n \iint |\hat{\sigma}(u,v,X_t)-\hat{\sigma}(u,v)|^2 a(X_t)dW(u,v) = h^{k/2}\sum_{t=1}^n \gamma_1(X_t)a(X_t)$ where $\gamma_1(X_t)$ is defined as in Theorem 4.4.2.

$$\begin{aligned}
&\iint |\hat{\sigma}(u,v,X_t)-\hat{\sigma}(u,v)|^2 dW(u,v)\\
= &\iint |\hat{\sigma}(u,v,X_t)|^2 dW(u,v) + \iint |\hat{\sigma}(u,v)|^2 dW(u,v)\\
&-2\iint Re\left[\hat{\sigma}(u,v,X_t)\hat{\sigma}(u,v)^*\right] dW(u,v)\\
= &B_1 + B_2 - 2B_3.
\end{aligned}$$

Given $\hat{\sigma}(u,v,X_t) = \hat{\phi}_{yz}(u,v,X_t) - \hat{\phi}_y(u,X_t)\hat{\phi}_z(v,X_t)$,

$$B_1 = \iint \left\{|\hat{\phi}(u,v,X_t)|^2 + |\hat{\phi}_y(u,X_t)\hat{\phi}_z(v,X_t)|^2 - 2Re\left[\hat{\phi}(u,v,X_t)\hat{\phi}_y(u,X_t)^*\hat{\phi}_z(v,X_t)^*\right]\right\} dW(u,v).$$

By $\hat{\phi}_{yz}(u,v,X_t) = \sum_{s=1}^n e^{iuY_s+ivZ_s}\hat{H}_{st}$ and $\hat{\phi}_y(u,X_t)\hat{\phi}_z(v,X_t) = \sum_{s,r}^n e^{iuY_s+ivZ_r}\hat{H}_{st}\hat{H}_{rt}$, we have

$$\begin{aligned}
B_1 = &\iint \Bigg\{ \sum_{s=1}^n [\cos(uY_s+vZ_s) + i\sin(uY_s+vZ_s))(\cos(uY_r+vZ_r) - i\sin(uY_r+vZ_r)]\hat{H}_{st}\hat{H}_{rt}\\
&+\sum_{s,r,l,m}[\cos(uY_s+vZ_l) + i\sin(uY_s+vZ_l))(\cos(uY_r+vZ_m) - i\sin(uY_r+vZ_m)]\hat{H}_{st}\hat{H}_{rt}\hat{H}_{lt}\hat{H}_{mt}\\
&-2Re\left(\sum_{s,r,l}[\cos(uY_s+vZ_s) + i\sin(uY_s+vZ_s))(\cos(uY_r+vZ_l) - i\sin(uY_r+vZ_l)]\right)\hat{H}_{st}\hat{H}_{rt}\hat{H}_{lt}\Bigg\}\\
&\times dW(u,v)\\
= &\iint \Bigg\{ \sum_{s=1}^n [\cos(uY_s+vZ_s)\cos(uY_r+vZ_r) + \sin(uY_s+vZ_s)\sin(uY_r+vZ_r)]\hat{H}_{st}\hat{H}_{rt}\\
&+\sum_{s,r,l,m}[\cos(uY_s+vZ_l)\cos(uY_r+vZ_m) + \sin(uY_s+vZ_l)\sin(uY_r+vZ_m)]\hat{H}_{st}\hat{H}_{rt}\hat{H}_{lt}\hat{H}_{mt}\\
&-2\sum_{s,r,l}[\cos(uY_s+vZ_s)\cos(uY_r+vZ_l) + \sin(uY_s+vZ_s)\sin(uY_r+vZ_l)]\hat{H}_{st}\hat{H}_{rt}\hat{H}_{lt}\Bigg\} dW(u,v)
\end{aligned}$$

$$
\begin{aligned}
= & \iint\Bigg\{\sum_{s=1}^{n}\cos[u(Y_s-Y_r)+v(Z_s-Z_r)]\hat{H}_{st}\hat{H}_{rt}+\sum_{s,r,l,m}\cos[u(Y_s-Y_r)+v(Z_l-Z_m)]\hat{H}_{st}\hat{H}_{rt}\hat{H}_{lt}\hat{H}_{mt} \\
& -2\sum_{s,r,l}\cos[u(Y_s-Y_r)+v(Z_s-Z_l)]\hat{H}_{st}\hat{H}_{rt}\hat{H}_{lt}\Bigg\}dW(u,v) \\
= & \iint\Bigg\{\sum_{s=1}^{n}\cos[u(Y_s-Y_r)]\cos[v(Z_s-Z_r)]\hat{H}_{st}\hat{H}_{rt}+\sum_{s,r,l,m}\cos[u(Y_s-Y_r)]\cos[v(Z_l-Z_m)] \\
& \times\hat{H}_{st}\hat{H}_{rt}\hat{H}_{lt}\hat{H}_{mt}-2\sum_{s,r,l}\cos[u(Y_s-Y_r)]\cos[v(Z_s-Z_l)]\hat{H}_{st}\hat{H}_{rt}\hat{H}_{lt}\Bigg\}dW(u,v),
\end{aligned}
$$

where the last equality follows from the fact that the sin function is an odd function and the weighting function $W(u,v)$ is symmetric about the origin. By the following equality

$$
\cos\alpha\cos\beta = 1-(1-\cos\alpha)-(1-\cos\beta)+(1-\cos\alpha)(1-\cos\beta)
$$

we can further show

$$
\begin{aligned}
B_1 = & \iint\Bigg\{\sum_{s=1}^{n}(1-[1-\cos u(Y_s-Y_r)]-[1-\cos v(Z_s-Z_r)] \\
& +[1-\cos u(Y_s-Y_r)][1-\cos v(Z_s-Z_r)])\hat{H}_{st}\hat{H}_{rt} \\
& +\sum_{s,r,l,m}(1-[1-\cos u(Y_s-Y_l)]-[1-\cos v(Z_r-Z_m)] \\
& +[1-\cos u(Y_s-Y_l)][1-\cos v(Z_r-Z_m)])\hat{H}_{st}\hat{H}_{rt}\hat{H}_{lt}\hat{H}_{mt} \\
& -2\sum_{s,r,l}(1-[1-\cos u(Y_s-Y_r)]-[1-\cos v(Z_s-Z_l)] \\
& +[1-\cos u(Y_s-Y_r)][1-\cos v(Z_s-Z_l)])\,\hat{H}_{st}\hat{H}_{rt}\hat{H}_{lt}\Bigg\}dW(u,v) \\
= & \sum_{s,r}|Y_s-Y_r|_p\,|Z_s-Z_r|_q\,\hat{H}_{st}\hat{H}_{rt}+\sum_{s,r,l,m}|Y_s-Y_l|_p\,|Z_r-Z_m|_q\,\hat{H}_{st}\hat{H}_{rt}\hat{H}_{lt}\hat{H}_{mt} \\
& -2\sum_{s,r,l}|Y_s-Y_r|_p\,|Z_s-Z_l|_q\,\hat{H}_{st}\hat{H}_{rt}\hat{H}_{lt},
\end{aligned}
$$

where we use the condition that $\sum_{s=1}^{n}\hat{H}_{st}=1$. Denote $P_{sr}\equiv|Y_s-Y_r|_p$ and $Q_{sr}\equiv|Z_s-Z_r|_q$, then we have

$$
\begin{aligned}
B_1 = & \textstyle\sum_{s,r}P_{sr}Q_{sr}\hat{H}_{st}\hat{H}_{rt}+\sum_{s,r,l,m}P_{sl}Q_{rm}\hat{H}_{st}\hat{H}_{rt}\hat{H}_{lt}\hat{H}_{mt} \\
& -2\textstyle\sum_{s,r,l}P_{sr}Q_{sl}\hat{H}_{st}\hat{H}_{rt}\hat{H}_{lt}.
\end{aligned}
$$

Furthermore, it is easy to show that

$$\begin{aligned} B_1 &= \sum_{s,r}\left\{\left(P_{sr}-\sum_r P_{sr}\hat{H}_{rt}-\sum_s P_{sr}\hat{H}_{st}+\sum_{s,r} P_{sr}\hat{H}_{st}\hat{H}_{rt}\right)\right. \\ &\quad \left.\times\left(Q_{sr}-\sum_r Q_{sr}\hat{H}_{rt}-\sum_s Q_{sr}\hat{H}_{st}+\sum_{s,r} Q_{sr}\hat{H}_{st}\hat{H}_{rt}\right)\right\}\hat{H}_{st}\hat{H}_{rt}. \end{aligned}$$

For B_2 and B_3, it is easy to show that they are similar to B_1 with slightly differences.

$$\begin{aligned} B_2 &= \frac{1}{n}^2\sum_{s,r}\left\{\left(P_{sr}-\frac{1}{n}\sum_r P_{sr}-\frac{1}{n}\sum_s P_{sr}+\frac{1}{n}^2\sum_{s,r} P_{sr}\right)\right. \\ &\quad \left.\times\left(Q_{sr}-\frac{1}{n}\sum_r Q_{sr}-\frac{1}{n}\sum_s Q_{sr}+\frac{1}{n}^2\sum_{s,r} Q_{sr}\right)\right\}, \end{aligned}$$

and

$$\begin{aligned} B_3 &= \frac{1}{n}\sum_{s,r}\left\{\left(P_{sr}-\frac{1}{n}\sum_r P_{sr}-\sum_s P_{sr}\hat{H}_{st}+\frac{1}{n}\sum_{s,r} P_{sr}\hat{H}_{st}\right)\right. \\ &\quad \left.\times\left(Q_{sr}-\frac{1}{n}\sum_r Q_{sr}-\sum_s Q_{sr}\hat{H}_{st}+\frac{1}{n}\sum_{s,r} Q_{sr}\hat{H}_{st}\right)\right\}\hat{H}_{st} \end{aligned}$$

$$\begin{aligned} B_3 = &\quad 1/n\textstyle\sum_{s,r}\left\{\left(P_{sr}-1/n\sum_r P_{sr}-\sum_s P_{sr}\hat{H}_{st}+1/n\sum_{s,r} P_{sr}\hat{H}_{st}\right)\right. \\ &\quad \left.\times\left(Q_{sr}-1/n\textstyle\sum_r Q_{sr}-\sum_s Q_{sr}\hat{H}_{st}+1/n\sum_{s,r} Q_{sr}\hat{H}_{st}\right)\right\}\hat{H}_{st}. \end{aligned}$$

Following the notation defined in Section 4.4, we have shown that

$$\gamma_1(X_t)=\sum\nolimits_{s,r}\left(A_{sr}^{(1)}B_{sr}^{(1)}/n^2+A_{sr}^{(2)}B_{sr}^{(2)}\hat{H}_{st}\hat{H}_{rt}-2A_{sr}^{(3)}B_{sr}^{(3)}\hat{H}_{st}/n\right).$$

Next, let's work on $\gamma_3(X_s,X_r)$ first in the asymptotic variance since $\gamma_2(X_s)$ in the asymptotic mean is simply a special case of $\gamma_3(X_s,X_r)$ by letting $s=r$. We want to show that

$$\begin{aligned} \gamma_3(X_s,X_r) &= \iint Re[(\hat{\varepsilon}_{yz}(u,v,X_s)-\hat{\phi}_z(v,X_s)\hat{\varepsilon}_y(u,X_s)-\hat{\phi}_y(u,X_s)\hat{\varepsilon}_z(v,X_s)) \\ &\quad \times(\hat{\varepsilon}_{yz}(u,v,X_r)-\hat{\phi}_z(v,X_r)\hat{\varepsilon}_y(u,X_r)-\hat{\phi}_y(u,X_r)\hat{\varepsilon}_z(v,X_r))^*]dW(u,v) \end{aligned}$$

$$\equiv \iint \Upsilon(u, v, X_s, X_r) dW(u, v),$$

where we let

$$\begin{aligned}\Upsilon(u, v, X_s, X_r) &\equiv Re[(\hat{\varepsilon}_{yz}(u, v, X_s) - \hat{\phi}_z(v, X_s)\hat{\varepsilon}_y(u, X_s) - \hat{\phi}_y(u, X_s)\hat{\varepsilon}_z(v, X_s))(\hat{\varepsilon}_{yz}(u, v, X_r) \\ &\quad -\hat{\phi}_z(v, X_r)\hat{\varepsilon}_y(u, X_r) - \hat{\phi}_y(u, X_r)\hat{\varepsilon}_z(v, X_r))^*].\end{aligned}$$

Furthermore, we show

$$\begin{aligned}\Upsilon(u, v, X_s, X_r) &= Re\Big\{\hat{\varepsilon}_{yz}(u, v, X_s)\hat{\varepsilon}_{yz}(u, v, X_r)^* - \hat{\phi}_z(v, X_s)\hat{\varepsilon}_y(u, X_s)\hat{\varepsilon}_{yz}(u, v, X_r)^* \\ &\quad -\hat{\phi}_y(u, X_s)\hat{\varepsilon}_z(v, X_s)\hat{\varepsilon}_{yz}(u, v, X_r)^* - \hat{\varepsilon}_{yz}(u, v, X_s)\hat{\phi}_z(v, X_r)^*\hat{\varepsilon}_y(u, X_r)^* \\ &\quad +\hat{\phi}_z(v, X_s)\hat{\varepsilon}_y(u, X_s)\hat{\phi}_z(v, X_r)^*\hat{\varepsilon}_y(u, X_r)^* + \hat{\phi}_z(v, X_s)\hat{\varepsilon}_y(u, X_s)\hat{\phi}_y(u, X_r)^*\hat{\varepsilon}_z(v, X_r)^* \\ &\quad -\hat{\varepsilon}_{yz}(u, v, X_s)\hat{\phi}_y(u, X_r)^*\hat{\varepsilon}_z(v, X_r)^* + \hat{\phi}_z(v, X_s)\hat{\varepsilon}_y(u, X_s)\hat{\phi}_y(u, X_r)^*\hat{\varepsilon}_z(v, X_r)^* \\ &\quad +\hat{\phi}_y(u, X_s)\hat{\varepsilon}_z(v, X_s)\hat{\phi}_y(u, X_r)^*\hat{\varepsilon}_z(v, X_r)^*\Big\} \\ &= \sum_{i=1}^{9} J_i, \text{ say.}\end{aligned}$$

For J_1, we have

$$\begin{aligned}J_1 &= Re\left[\hat{\varepsilon}_{yz}(u, v, X_s)\hat{\varepsilon}_{yz}(u, v, X_r)^*\right] \\ &= Re\left[\left(e^{iuY_s+ivZ_s} - \sum_{l=1}^{n} e^{iuY_l+ivZ_l}\hat{H}_{ls}\right)\left(e^{iuY_r+ivZ_r} - \sum_{m=1}^{n} e^{iuY_m+ivZ_m}\hat{H}_{mr}\right)^*\right] \\ &= Re\Bigg\{e^{iu(Y_s-Y_r)+iv(Z_s-Z_r)} + \sum_{l,m} e^{iu(Y_l-Y_m)+iv(Z_l-Z_m)}\hat{H}_{ls}\hat{H}_{mr} \\ &\quad - \sum_{m=1}^{n} e^{iu(Y_s-Y_m)+iv(Z_s-Z_m)}\hat{H}_{mr} - \sum_{l=1}^{n} e^{iu(Y_l-Y_r)+iv(Z_l-Z_r)}\hat{H}_{ls}\Bigg\} \\ &= [\cos u(Y_s - Y_r)]\,[\cos v(Z_s - Z_r)] + \sum_{l,m} [\cos u(Y_l - Y_m)]\,[\cos v(Z_l - Z_m)]\,\hat{H}_{ls}\hat{H}_{mr} \\ &\quad - \sum_{m=1}^{n} [\cos u(Y_s - Y_m)]\,[\cos v(Z_s - Z_m)]\,\hat{H}_{ms} - \sum_{l=1}^{n} [\cos u(Y_l - Y_r)]\,[\cos v(Z_l - Z_r)]\,\hat{H}_{ls}\end{aligned}$$

Applying the special weighting function that $dW(u, v) = \frac{1}{c_p|u|_p^{p+1}}\frac{1}{c_q|v|_q^{q+1}}dudv$, where $c_p = \frac{\pi^{(1+p)/2}}{\Gamma((1+p)/2)}$ and $c_q = \frac{\pi^{(1+q)/2}}{\Gamma((1+q)/2)}$ and by Lemma (Székely et al. 2007), we show

$$\iint J_1 dW(u, v) = P_{sr}Q_{sr} - P_{sr} - Q_{sr}$$

$$
\begin{aligned}
&+\sum_{l,m} P_{lm}Q_{lm}\hat{H}_{ls}\hat{H}_{mr}-\sum_{l,m} P_{lm}\hat{H}_{ls}\hat{H}_{mr}-\sum_{l,m} Q_{lm}\hat{H}_{ls}\hat{H}_{mr}\\
&+\sum_{l} P_{rl}\hat{H}_{ls}+\sum_{l} Q_{rl}\hat{H}_{ls}-\sum_{l} P_{rl}Q_{rl}\hat{H}_{ls}\\
&+\sum_{m} P_{sm}\hat{H}_{mr}+\sum_{m} Q_{sm}\hat{H}_{mr}-\sum_{m} P_{sm}Q_{sm}\hat{H}_{mr}.
\end{aligned}
$$

By the similar steps, we can show

$$
\begin{aligned}
\iint J_2 dW(u,v) \;=\; & -P_{sr}\sum_{l} Q_{rl}\hat{H}_{ls}+P_{sr}+\sum_{l} Q_{rl}\hat{H}_{ls}\\
& -\sum_{l,m,w} P_{lw}Q_{mw}\hat{H}_{ls}\hat{H}_{ms}\hat{H}_{wr}+\sum_{l,m} P_{lm}\hat{H}_{ls}\hat{H}_{mr}+\sum_{l,m} Q_{lm}\hat{H}_{ls}\hat{H}_{mr}\\
& -\sum_{l} P_{rl}\hat{H}_{ls}-\sum_{l} Q_{rl}\hat{H}_{ls}+\sum_{l,m} P_{rl}Q_{rm}\hat{H}_{ls}\hat{H}_{ms}\\
& -\sum_{m} P_{sm}\hat{H}_{mr}-\sum_{l,m} Q_{lm}\hat{H}_{ls}\hat{H}_{mr}+\sum_{l,m} P_{sm}Q_{lm}\hat{H}_{ls}\hat{H}_{mr},
\end{aligned}
$$

$$
\begin{aligned}
\iint J_3 dW(u,v) \;=\; & -\sum_{l} P_{rl}Q_{rs}\hat{H}_{ls}+Q_{sr}+\sum_{l} P_{rl}\hat{H}_{ls}\\
& -\sum_{l,m,w} P_{lw}Q_{mw}\hat{H}_{ls}\hat{H}_{ms}\hat{H}_{wr}+\sum_{l,m} P_{lm}\hat{H}_{ls}\hat{H}_{mr}+\sum_{l,m} Q_{lm}\hat{H}_{ls}\hat{H}_{mr}\\
& -\sum_{l} P_{rl}\hat{H}_{ls}-\sum_{l} Q_{rl}\hat{H}_{ls}+\sum_{l,m} P_{rl}Q_{rm}\hat{H}_{ls}\hat{H}_{ms}\\
& -\sum_{l,m} P_{lm}\hat{H}_{ls}\hat{H}_{mr}-\sum_{l} Q_{sl}\hat{H}_{lr}+\sum_{l,m} P_{lm}Q_{sm}\hat{H}_{ls}\hat{H}_{mr},
\end{aligned}
$$

$$
\begin{aligned}
\iint J_4 dW(u,v) \;=\; & -P_{sr}\sum_{l} Q_{sl}\hat{H}_{lr}+P_{sr}+\sum_{l} Q_{sl}\hat{H}_{lr}\\
& -\sum_{l,m,w} P_{lm}Q_{lw}\hat{H}_{ls}\hat{H}_{mr}\hat{H}_{wr}+\sum_{l,m} P_{lm}\hat{H}_{ls}\hat{H}_{mr}+\sum_{l,m} Q_{lm}\hat{H}_{ls}\hat{H}_{mr}\\
& -\sum_{l} P_{rl}\hat{H}_{ls}-\sum_{l,m} Q_{lm}\hat{H}_{ls}\hat{H}_{mr}+\sum_{l,m} P_{rl}Q_{lm}\hat{H}_{ls}\hat{H}_{mr}\\
& -\sum_{l} P_{sl}\hat{H}_{lr}-\sum_{l} Q_{sl}\hat{H}_{lr}+\sum_{l,m} P_{sl}Q_{sm}\hat{H}_{lr}\hat{H}_{mr},
\end{aligned}
$$

$$
\begin{aligned}
\iint J_5 dW(u,v) \;=\; & P_{sr}\sum_{l,m} Q_{lm}\hat{H}_{ls}\hat{H}_{mr}-P_{sr}-\sum_{l,m} Q_{lm}\hat{H}_{ls}\hat{H}_{mr}\\
& +\sum_{l,m,w,j} P_{lw}Q_{mj}\hat{H}_{ls}\hat{H}_{ms}\hat{H}_{wr}\hat{H}_{jr}-\sum_{l,m} P_{lm}\hat{H}_{ls}\hat{H}_{mr}-\sum_{l,m} Q_{lm}\hat{H}_{ls}\hat{H}_{mr}
\end{aligned}
$$

$$+\sum_{l} P_{rl}\hat{H}_{ls} + \sum_{l,m} Q_{lm}\hat{H}_{ls}\hat{H}_{mr} - \sum_{l,m,w} P_{rl}Q_{mw}\hat{H}_{ls}\hat{H}_{ms}\hat{H}_{mr}$$

$$+\sum_{l} P_{sl}\hat{H}_{lr} + \sum_{l,m} Q_{lm}\hat{H}_{ls}\hat{H}_{mr} - \sum_{l,m,w} P_{sl}Q_{wm}\hat{H}_{lr}\hat{H}_{ww}\hat{H}_{mr},$$

$$\begin{aligned}
\iint J_6 dW(u,v) \;=\; & -\sum_{l} P_{rl}\hat{H}_{ls} - \sum_{l} Q_{sl}\hat{H}_{lr} + \sum_{l,m} P_{rl}Q_{sm}\hat{H}_{ls}\hat{H}_{mr} \\
& +\sum_{l} P_{rl}\hat{H}_{ls} + \sum_{l,m} Q_{lm}\hat{H}_{ls}\hat{H}_{mr} - \sum_{l,m,w} P_{rl}Q_{mw}\hat{H}_{ls}\hat{H}_{ms}\hat{H}_{wr} \\
& +\sum_{l,m} P_{lm}\hat{H}_{ls}\hat{H}_{mr} + \sum_{l} Q_{sl}\hat{H}_{lr} - \sum_{l,m,w} P_{lm}Q_{sw}\hat{H}_{ls}\hat{H}_{mr}\hat{H}_{wr} \\
& -\sum_{l,m} P_{lm}\hat{H}_{ls}\hat{H}_{mr} - \sum_{l,m} Q_{lm}\hat{H}_{ls}\hat{H}_{mr} + \sum_{l,m,w,j} P_{lw}Q_{mj}\hat{H}_{ls}\hat{H}_{ms}\hat{H}_{wr}\hat{H}_{jr},
\end{aligned}$$

$$\begin{aligned}
\iint J_7 dW(u,v) \;=\; & \sum_{l} P_{sl}\hat{H}_{lr} + Q_{sr} - \sum_{l} P_{sl}Q_{sr}\hat{H}_{lr} \\
& -\sum_{l,m} P_{lm}\hat{H}_{ls}\hat{H}_{mr} - \sum_{l} Q_{rl}\hat{H}_{ls} + \sum_{l,m} P_{lm}Q_{rl}\hat{H}_{ls}\hat{H}_{mr} \\
& -\sum_{l} P_{sl}\hat{H}_{lr} - \sum_{l} Q_{sl}\hat{H}_{lr} + \sum_{l,m} P_{sl}Q_{sm}\hat{H}_{lr}\hat{H}_{mr} \\
& +\sum_{l,m} P_{lm}\hat{H}_{ls}\hat{H}_{mr} + \sum_{l,m} Q_{lm}\hat{H}_{ls}\hat{H}_{mr} - \sum_{l,m,w} P_{lm}Q_{lw}\hat{H}_{ls}\hat{H}_{mr}\hat{H}_{wr},
\end{aligned}$$

$$\begin{aligned}
\iint J_8 dW(u,v) \;=\; & -\sum_{l} P_{sl}\hat{H}_{lr} - \sum_{l} Q_{lr}\hat{H}_{ls} + \sum_{l,m} P_{sm}Q_{lr}\hat{H}_{ls}\hat{H}_{mr} \\
& +\sum_{l,m} P_{lm}\hat{H}_{ls}\hat{H}_{mr} + \sum_{l} Q_{lr}\hat{H}_{ls} - \sum_{l,m,w} P_{lw}Q_{mr}\hat{H}_{ls}\hat{H}_{ms}\hat{H}_{wr} \\
& +\sum_{l} P_{sl}\hat{H}_{lr} + \sum_{l,m} Q_{lm}\hat{H}_{ls}\hat{H}_{mr} - \sum_{l,m,w} P_{sm}Q_{lw}\hat{H}_{ls}\hat{H}_{mr}\hat{H}_{wr} \\
& -\sum_{l,m} P_{lm}\hat{H}_{ls}\hat{H}_{mr} - \sum_{l,m} Q_{lm}\hat{H}_{ls}\hat{H}_{mr} + \sum_{l,m,w,j} P_{lw}Q_{mj}\hat{H}_{ls}\hat{H}_{ms}\hat{H}_{wr}, \hat{H}_{jr},
\end{aligned}$$

and

$$\begin{aligned}
\iint J_9 dW(u,v) \;=\; & -\sum_{l,m} P_{lm}\hat{H}_{lr}\hat{H}_{mr} - Q_{sr} + Q_{sr}\sum_{l,m} P_{lm}\hat{H}_{ls}\hat{H}_{mr} \\
& +\sum_{l,m} P_{lm}\hat{H}_{ls}\hat{H}_{mr} + \sum_{l} Q_{lr}\hat{H}_{ls} - \sum_{l,m,w} P_{lw}Q_{mr}\hat{H}_{ls}\hat{H}_{ms}\hat{H}_{wr} \\
& +\sum_{l,m} P_{lm}\hat{H}_{ls}\hat{H}_{mr} + \sum_{l} Q_{sl}\hat{H}_{lr} - \sum_{l,m,w} P_{lm}Q_{sw}\hat{H}_{ls}\hat{H}_{mr}\hat{H}_{wr}
\end{aligned}$$

$$-\sum_{l,m} P_{lm}\hat{H}_{ls}\hat{H}_{mr} - \sum_{l,m} Q_{lm}\hat{H}_{ls}\hat{H}_{mr} + \sum_{l,m,w,j} P_{lw}Q_{mj}\hat{H}_{ls}\hat{H}_{ms}\hat{H}_{wr}\hat{H}_{jr}.$$

Combining the $\iint J_1 dW(u,v)$, $\iint J_2 dW(u,v)$, ..., and $\iint J_9 dW(u,v)$ we have proved

$$\hat{V}_n^F = \hat{V}_n^W \equiv 2h^{k/2} \sum_{1\leq s<t\leq n} \left[\sum_{t=1}^{n} a(X_t)\hat{H}_{st}\hat{H}_{rt}\gamma_3(X_s, X_r) \right]^2$$

At last, for the integration in finite sample mean, we just need to let $s = r$ in $J_1, ..., J_9$ and the result follows immediately. Therefore, we have proved Theorem 4.4.2. □

BIBLIOGRAPHY

[1] Chunrong Ai and Xiaohong Chen. Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica*, 71(6):1795–1843, 2003.

[2] Donald WK Andrews. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica: Journal of the Econometric Society*, pages 817–858, 1991.

[3] Donald WK Andrews. Tests for parameter instability and structural change with unknown change point. *Econometrica: Journal of the Econometric Society*, pages 821–856, 1993.

[4] Donald WK Andrews and Ray C Fair. Inference in nonlinear econometric models with structural change. *The Review of Economic Studies*, 55(4):615–640, 1988.

[5] Donald WK Andrews and Werner Ploberger. Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica: Journal of the Econometric Society*, pages 1383–1414, 1994.

[6] Andrew Ang and Geert Bekaert. Regime switches in interest rates. *Journal of Business & Economic Statistics*, 20(2):163–182, 2002.

[7] Andrew Ang and Joseph Chen. Asymmetric correlations of equity portfolios. *Journal of financial Economics*, 63(3):443–494, 2002.

[8] Jushan Bai and Serena Ng. Forecasting economic time series using targeted predictors. *Journal of Econometrics*, 146(2):304–317, 2008.

[9] Jushan Bai and Pierre Perron. Estimating and testing linear models with multiple structural changes. *Econometrica*, pages 47–78, 1998.

[10] Jushan Bai and Pierre Perron. Computation and analysis of multiple structural change models. *Journal of applied econometrics*, 18(1):1–22, 2003.

[11] Kent G Becker, Joseph E Finnerty, and Manoj Gupta. The intertemporal relation between the us and japanese stock markets. *The Journal of Finance*, 45(4):1297–1306, 1990.

[12] Herman J Bierens. Consistent model specification tests. *Journal of Econometrics*, 20(1):105–134, 1982.

[13] Monica Billio, Mila Getmansky, Andrew W Lo, and Loriana Pelizzon. Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of Financial Economics*, 104(3):535–559, 2012.

[14] Richard Blundell, Xiaohong Chen, and Dennis Kristensen. Semi-nonparametric iv estimation of shape-invariant engel curves. *Econometrica*, 75(6):1613–1669, 2007.

[15] Richard C Bradley et al. Basic properties of strong mixing conditions. a survey and some open questions. *Probability surveys*, 2(2):107–144, 2005.

[16] Richard C Bradley and Cristina Tone. A central limit theorem for non-stationary strongly mixing random fields. *Journal of Theoretical Probability*, pages 1–20, 2015.

[17] Zongwu Cai. Trending time-varying coefficient time series models with serially correlated errors. *Journal of Econometrics*, 136(1):163–188, 2007.

[18] Zongwu Cai, Jianqing Fan, and Qiwei Yao. Functional-coefficient regression models for nonlinear time series. *Journal of the American Statistical Association*, 95(451):941–956, 2000.

[19] Zongwu Cai, Yunfei Wang, and Yonggang Wang. Testing instability in a predictive regression model with nonstationary regressors. *Econometric Theory*, FirstView:1–28, 12 2014.

[20] Zongwu Cai, Yunfei Wang, and Yonggang Wang. Testing instability in a predictive regression model with nonstationary regressors. *Econometric Theory*, 31(05):953–980, 2015.

[21] Guillermo A Calvo. Staggered prices in a utility-maximizing framework. *Journal of monetary Economics*, 12(3):383–398, 1983.

[22] John Y Campbell and Samuel B Thompson. Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies*, 21(4):1509–1531, 2008.

[23] Bin Chen. Modeling and testing smooth structural changes with endogenous regressors. *Journal of Econometrics*, 185(1):196–215, 2015.

[24] Bin Chen and Yongmiao Hong. Characteristic function–based testing for multifactor continuous-time markov models via nonparametric regression. *Econometric Theory*, 26(04):1115–1179, 2010.

[25] Bin Chen and Yongmiao Hong. Testing for smooth structural changes in time series models via nonparametric regression. *Econometrica*, 80(3):1157–1183, 2012.

[26] Bin Chen and Yongmiao Hong. Detecting for smooth structural changes in garch models. *Econometric Theory*, 32(03):740–791, 2016.

[27] In Choi and Peter CB Phillips. Testing for a unit root by frequency domain regression. *Journal of Econometrics*, 59(3):263–286, 1993.

[28] Gregory C Chow. Tests of equality between sets of coefficients in two linear regressions. *Econometrica: Journal of the Econometric Society*, pages 591–605, 1960.

[29] Michael P Clements, Philip Hans Franses, and Norman R Swanson. Forecasting economic and financial time-series with non-linear models. *International Journal of Forecasting*, 20(2):169–183, 2004.

[30] Robert A Connolly and F Albert Wang. International equity market comovements: Economic fundamentals or contagion? *Pacific-Basin Finance Journal*, 11(1):23–43, 2003.

[31] Dean Corbae, Sam Ouliaris, and Peter CB Phillips. Band spectral regression with trending data. *Econometrica*, 70(3):1067–1109, 2002.

[32] Alastair Craig, Ajay Dravid, and Matthew Richardson. Market efficiency around the clock some supporting evidence using foreign-based derivatives. *Journal of Financial Economics*, 39(2):161–180, 1995.

[33] Rainer Dahlhaus. On the kullback-leibler information divergence of locally stationary processes. *Stochastic Processes and their Applications*, 62(1):139–168, 1996.

[34] Rainer Dahlhaus. Locally stationary processes. *Handbook of Statistics, Time Series Analysis: Methods and Applications*, pages 351–408, 2012.

[35] Serge Darolles, Yanqin Fan, Jean-Pierre Florens, and Eric Renault. Nonparametric instrumental regression. *Econometrica*, 79(5):1541–1565, 2011.

[36] Jan G De Gooijer and Kuldeep Kumar. Some recent developments in nonlinear time series modelling, testing, and forecasting. *International Journal of Forecasting*, 8(2):135–156, 1992.

[37] Graham Elliott and Ulrich K Müller. Efficient tests for general persistent time variation in regression coefficients. *The Review of Economic Studies*, 73(4):907–940, 2006.

[38] Robert Engle. Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, 20(3):339–350, 2002.

[39] Robert F Engle. Band spectrum regression. *International Economic Review*, pages 1–11, 1974.

[40] Arturo Estrella and Jeffrey C Fuhrer. Dynamic inconsistencies: Counterfactual implications of a class of rational-expectations models. *The American Economic Review*, 92(4):1013–1028, 2002.

[41] Arturo Estrella and Jeffrey C Fuhrer. Monetary policy shifts and the stability of monetary policy models. *Review of Economics and Statistics*, 85(1):94–104, 2003.

[42] J Fan and Irène Gijbels. Local polynomial modelling and its applications. 1996.

[43] Jianqing Fan and Qiwei Yao. Nonlinear time series: nonparametric and parametric methods. 2003.

[44] Kristin J Forbes and Roberto Rigobon. No contagion, only interdependence: measuring stock market comovements. *The journal of Finance*, 57(5):2223–2261, 2002.

[45] Jeff Fuhrer and George Moore. Inflation persistence. *The Quarterly Journal of Economics*, pages 127–159, 1995.

[46] Jeffrey C Fuhrer. The (un) importance of forward-looking behavior in price specifications. *Journal of Money, Credit, and Banking*, pages 338–350, 1997.

[47] Jordi Gali and Mark Gertler. Inflation dynamics: A structural econometric analysis. *Journal of monetary Economics*, 44(2):195–222, 1999.

[48] Jordi Gali, Mark Gertler, and J David Lopez-Salido. Robustness of the estimates of the hybrid new keynesian phillips curve. *Journal of Monetary Economics*, 52(6):1107–1118, 2005.

[49] Gloria González'Rivera and TH Lee. Nonlinear time series in financial forecasting. *Encyclopedia of Complexity and Systems Science. Springer, New York*, 2009.

[50] Clive WJ Granger and Mark W Watson. Time series and spectral methods in econometrics. *Handbook of econometrics*, 2:979–1022, 1984.

[51] CWJ Granger and M Hatanaka. Spectral analysis of economic time series. 1964.

[52] Herbert G Grubel and Kenneth Fadner. The interdependence of international equity markets. *The Journal of Finance*, 26(1):89–94, 1971.

[53] Alastair R Hall, Sanggohn Han, and Otilia Boldea. Inference regarding multiple structural changes in linear models with endogenous regressors. *Journal of econometrics*, 170(2):281–302, 2012.

[54] Peter Hall, Joel L Horowitz, et al. Nonparametric methods for inference in the presence of instrumental variables. *The Annals of Statistics*, 33(6):2904–2929, 2005.

[55] Peter Hall and Thomas E Wehrly. A geometrical method for removing edge effects from kernel-type nonparametric regression estimators. *Journal of the American Statistical Association*, 86(415):665–672, 1991.

[56] James D Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the Econometric Society*, pages 357–384, 1989.

[57] Edward J Hannan. The estimation of relationships involving distributed lags. *Econometrica: Journal of the Econometric Society*, pages 206–224, 1965.

[58] Edward J Hannan. The estimation of a lagged regression relation. *Biometrika*, 54(3-4):409–418, 1967.

[59] Bruce E Hansen. Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica: Journal of the econometric society*, pages 413–430, 1996.

[60] Bruce E Hansen. Testing for structural change in conditional models. *Journal of Econometrics*, 97(1):93–115, 2000.

[61] Bruce E Hansen. The new econometrics of structural change: Dating breaks in us labor productivity. *Journal of Economic perspectives*, pages 117–128, 2001.

[62] Jerry A Hausman. Specification tests in econometrics. *Econometrica: Journal of the Econometric Society*, pages 1251–1271, 1978.

[63] Javier Hidalgo. A nonparametric conditional moment test for structural stability. *Econometric Theory*, 11:671–671, 1995.

[64] Eric Hillebrand, Tae&Hwy Lee, and M Medeiros. Bagging constrained forecasts with application to forecasting equity premium. *JSM Proceedings for Business and Economic Statistics*, 2009.

[65] Vidar Hjellvik, Qiwei Yao, and Dag Tjøstheim. Linearity testing using local polynomial approximation. *Journal of Statistical Planning and Inference*, 68(2):295–321, 1998.

[66] Yongmiao Hong. Hypothesis testing in time series via the empirical characteristic function: a generalized spectral density approach. *Journal of the American Statistical Association*, 94(448):1201–1220, 1999.

[67] Yongmiao Hong, Jun Tu, and Guofu Zhou. Asymmetries in stock returns: Statistical tests and economic evaluation. *Review of Financial Studies*, 20(5):1547–1581, 2007.

[68] Yongmiao Hong, Shouyang Wang, and Xia Wang. Testing strict stationarity with applications to macroeconomic and financial time series. *Working paper*, 2015.

[69] Joel L Horowitz. Applied nonparametric instrumental variables estimation. *Econometrica*, 79(2):347–394, 2011.

[70] Alan Jeffrey and Daniel Zwillinger. *Table of integrals, series, and products*. Academic Press, 2007.

[71] G Andrew Karolyi and René M Stulz. Why do markets move together? an investigation of us-japan stock return comovements. *The Journal of Finance*, 51(3):951–986, 1996.

[72] Dennis Kristensen. Non-parametric detection and estimation of structural change. *The Econometrics Journal*, 15(3):420–461, 2012.

[73] Soumendra N Lahiri. Theoretical comparisons of block bootstrap methods. *Annals of Statistics*, pages 386–404, 1999.

[74] Tae-Hwy Lee, Halbert White, and Clive WJ Granger. Testing for neglected nonlinearity in time series models: A comparison of neural network methods and alternative tests. *Journal of Econometrics*, 56(3):269–290, 1993.

[75] Luiz Renato Lima and Zhijie Xiao. Is there long memory in financial time series? *Applied Financial Economics*, 20(6):487–500, 2010.

[76] Chien-Fu Jeff Lin and Timo Teräsvirta. Testing the constancy of regression parameters against continuous structural change. *Journal of Econometrics*, 62(2):211–228, 1994.

[77] Wen-Ling Lin, Robert F Engle, and Takatoshi Ito. Do bulls and bears move across borders? international transmission of stock returns and volatility. *Review of financial studies*, 7(3):507–538, 1994.

[78] Francois Longin and Bruno Solnik. Is the correlation in international equity returns constant: 1960–1990? *Journal of international money and finance*, 14(1):3–26, 1995.

[79] Francois Longin and Bruno Solnik. Extreme correlation of international equity markets. *The journal of finance*, 56(2):649–676, 2001.

[80] Whitney K Newey and James L Powell. Instrumental variable estimation of nonparametric models. *Econometrica*, 71(5):1565–1578, 2003.

[81] Whitney K Newey, James L Powell, and Francis Vella. Nonparametric estimation of triangular simultaneous equations models. *Econometrica*, 67(3):565–603, 1999.

[82] Whitney K Newey and Kenneth D West. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708, 1987.

[83] Susan Orbe, Eva Ferreira, and Juan Rodríguez-póo. A nonparametric

method to estimate time varying coefficients under seasonal constraints. *Journal of nonparametric statistics*, 12(6):779–806, 2000.

[84] Susan Orbe, Eva Ferreira, and Juan Rodríguez-póo. Nonparametric estimation of time varying parameters under shape restrictions. *Journal of Econometrics*, 126(1):53–77, 2005.

[85] Andrew J Patton. Modelling asymmetric exchange rate dependence. *International economic review*, 47(2):527–556, 2006.

[86] Pierre Perron. Dealing with structural breaks. *Palgrave handbook of econometrics*, 1:278–352, 2006.

[87] Pierre Perron and Zhongjun Qu. Estimating restricted structural change models. *Journal of Econometrics*, 134(2):373–399, 2006.

[88] Pierre Perron and Yohei Yamamoto. A note on estimating and testing for multiple structural changes in models with endogenous regressors via 2sls. *Econometric Theory*, 30(02):491–507, 2014.

[89] Pierre Perron and Yohei Yamamoto. Using ols to estimate and test for structural changes in models with endogenous regressors. *Journal of Applied Econometrics*, 30(1):119–144, 2015.

[90] Dimitris N Politis and Halbert White. Automatic block-length selection for the dependent bootstrap. *Econometric Reviews*, 23(1):53–70, 2004.

[91] Richard E Quandt. Tests of the hypothesis that a linear regression system obeys two separate regimes. *Journal of the American statistical Association*, 55(290):324–330, 1960.

[92] Peter M Robinson. Nonparametric estimation of time-varying parameters. In *Statistical Analysis and Forecasting of Economic Structural Change*, pages 253–264. Springer, 1989.

[93] Peter M Robinson. *Time-varying nonlinear regression*. Springer, 1991.

[94] Murray Rosenblatt. A quadratic measure of deviation of two-dimensional density estimates and a test of independence. *The Annals of Statistics*, pages 1–14, 1975.

[95] George G Roussas and D Ioannides. Moment inequalities for mixing sequences of random variables. *Stochastic Analysis and Applications*, 5(1):60–120, 1987.

[96] Glenn D Rudebusch. Assessing nominal income rules for monetary policy with model and data uncertainty. *The Economic Journal*, 112(479):402–432, 2002.

[97] Argia M Sbordone. Prices and unit labor costs: a new test of price stickiness. *Journal of Monetary economics*, 49(2):265–292, 2002.

[98] Argia M Sbordone. Do expected future marginal costs drive inflation dynamics? *Journal of Monetary Economics*, 52(6):1183–1197, 2005.

[99] James H Stock and Mark W Watson. Evidence on structural instability in macroeconomic time series relations. *Journal of Business & Economic Statistics*, 14(1):11–30, 1996.

[100] Jim Stock and Mark Watson. Forecasting in dynamic factor models subject to structural instability. *The Methodology and Practice of Econometrics. A Festschrift in Honour of David F. Hendry*, 173:205, 2009.

[101] Liangjun Su and Halbert White. A consistent characteristic function-based test for conditional independence. *Journal of Econometrics*, 141(2):807–834, 2007.

[102] Liangjun Su and Halbert White. A nonparametric hellinger metric test for conditional independence. *Econometric Theory*, 24(04):829–864, 2008.

[103] Liangjun Su and Halbert White. Testing structural change in partially linear models. *Econometric Theory*, 26(06):1761–1806, 2010.

[104] Liangjun Su and Zhijie Xiao. Testing structural change in time-series nonparametric regression models. *Statistics and its Interface*, 1:347–366, 2008.

[105] Gábor J Székely, Maria L Rizzo, Nail K Bakirov, et al. Measuring and testing dependence by correlation of distances. *The Annals of Statistics*, 35(6):2769–2794, 2007.

[106] Carlos Tenreiro. Loi asymptotique des erreurs quadratiques intégrées des estimateursa noyau de la densité et de la régression sous des conditions de dépendance.

[107] Timo Teräsvirta. Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the american Statistical association*, 89(425):208–218, 1994.

[108] Timo Teräsvirta, Dag Tjøstheim, and Clive William John Granger. *Modelling nonlinear economic time series*. Oxford University Press Oxford, 2010.

[109] Panayiotis Theodossiou and Unro Lee. Mean and volatility spillovers across major national stock markets: Further empirical evidence. *Journal of Financial Research*, 16(4):337–350, 1993.

[110] H Tong. Threshold models in nonlinear time series analysis. 1983.

[111] Michael Vogt. Testing for structural change in time-varying nonparametric regression models. *Econometric Theory*, 31(04):811–859, 2015.

[112] X Wang and Y Hong. Characteristic function based testing for conditional independence: A nonparametric regression approach. Technical report, Working paper, Xiamen University, 2012.

[113] Ivo Welch and Amit Goyal. A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, 21(4):1455–1508, 2008.

[114] Halbert White. Some asymptotic results for learning in single hidden-layer feedforward network models. *Journal of the American Statistical Association*, 84(408):1003–1013, 1989.

[115] Ke-Li Xu. Testing for structural change under non-stationary variances. *The Econometrics Journal*, 18(2):274–305, 2015.

[116] Xia Wang Yongmiao Hong and Shouyang Wang. Testing strict stationrity with applications to macroeconomic and financial time series. *Working paper, Xiamen University*, 2014.

[117] Ken-ichi Yoshihara. Limiting behavior of u-statistics for stationary, absolutely regular processes. *Probability Theory and Related Fields*, 35(3):237–252, 1976.

[118] Chengsi Zhang, Denise R Osborn, and Dong Heon Kim. The new keynesian phillips curve: from sticky inflation to sticky prices. *Journal of Money, Credit and Banking*, 40(4):667–699, 2008.

[119] Ting Zhang and Wei Biao Wu. Inference of time-varying regression models. *The Annals of Statistics*, 40(3):1376–1402, 2012.

[120] Zhou Zhou. Measuring nonlinear dependence in time-series, a distance correlation approach. *Journal of Time Series Analysis*, 33(3):438–457, 2012.